REMARKS/ARGUMENTS

This amendment is filed in response to the Office Action mailed April 6, 2010 for the above captioned application. Reconsideration of the application as amended in view of the remarks herein is respectfully requested.

Applicants request an extension of time sufficient to make this paper timely and enclose the appropriate fee.

Applicants thank the Examiner for the clear statement of the status of the prior rejections.

Anticipation Rejection

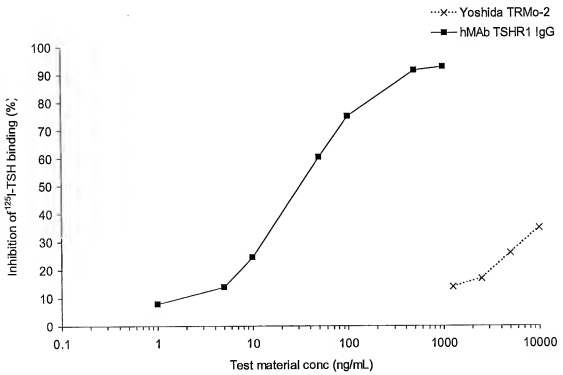
Claims 121, 127, 127, 129, 130, 133, 136, 137 and 198 stand rejected as anticipated by Yoshida et al. The Examiner contends that Yoshida discloses a monoclonal antibody TrMo-2 which falls within the scope of the claims, and transfers the burden to Applicants to test this antibody to show whether the characteristics are accurately reported because the USPTO lacks testing facilities. Applicants respectfully traverse this rejection on several grounds.

First, Applicants submit that Yoshida et al. fails to provide an enabling disclosure of TRMo-2. The formation of any given monoclonal antibody is a matter of random chance. The specific monoclonal antibody does not appear to have been deposited with a repository of biological materials such as the American Type Culture Collection (ATCC) nor it is characterized by sequence or in some other manner that would allow a newly formed antibody to be identified as being the same as TRMo-2. This lack of an enabled source for this specific antibody makes it impossible for Applicants to test its properties or compare these properties directly with the antibodies they have developed. Since a disclosure must be enabling in order to be anticipatory, reliance on TRMo-2 as mentioned in Yoshida as evidence of anticipation is improper. The rejection as to all claims should therefore be withdrawn.

With respect to claim 198, the Examiner does not provide any specific citation to a teaching of the recited level of affinity within the Yoshida reference. Yoshida does not provide a numeric affinity of TRMo-2 for the TSH receptor. It can be pointed out, however, that the numeric results that are provided are very different from those of the present invention. For example, the binding studies reported in Fig. 1 of Yoshida use antibody concentrations in the range of 1.25 to 10 μg/ml. In contrast, Fig. 1 of the present application shows inhibition of TSH binding at concentrations around 10-100 ng/ml This indicates several orders of magnitude greater affinity for the receptor. This can be seen from the following Table and Figure which summarize the data from the two references in one place.

Table 1

	Inhibition of ^{125}I-TSH binding to the TSHR (%)	
Antibody concentration(ng/mL)	**Yoshida TRMo-2**	**hMAb TSHR1 IgG**
0.5		0.1
1		8
5		14
10		24.7
50		60.6
100		75.1
500		91.6
1000		92.8
1250	14	
2500	17	
500	26	
10000	35	



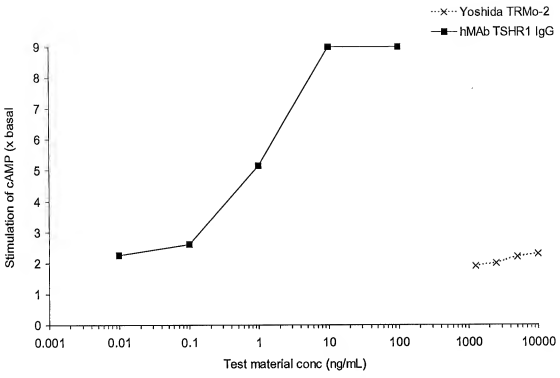
The same is true of the quantitative limits set in claims 126, 127, 129, 130, 133, 136 and 137. The Examiner has incorrectly referred to the units of those claims as "internal" units, when in fact these are units of an international standard. NIBSC refers to the National Institute for Biological Standards and Controls, which is a standards agency of the United Kingdom and a WHO International Laboratory for Biological Standards. Standard 90/672 is a preparation available from NIBSC that exhibits a standard specific activity to which the specific activity of an sample can be compared. (See exhibit A).

In terms of inhibition of TSH binding (Table 1 and Figure 1, above), hMAb TSHR1 is approximately 300x more potent than TRMo-2. hMAb TSHR1 has a specific activity of 150 units per mg (Table 7 in the specification), thus TRMo-2 has a specific activity of about 0.5 units per mg in terms of inhibition of TSH binding which is well outside any of the numeric ranges claimed.

A similar comparison to that made above can be made for cAMP stimulation.

Table 2 Stimulation of cyclic AMP production

	Stimulation of cyclic AMP production (x basal)	
Antibody concentration ng/mL	**Yoshida TRMo-2**	**hMAb TSHR1 IgG**
0.01		2.26
0.02		
0.1		2.61
0.2		
1		5.16
2		
10		9
20		
100		9
1250	1.9	
2500	2	
5000	2.2	
10000	2.3	



hMAb TSHR1 has a specific activity of 318 units per mg (Table 8 in the specification). As shown in Fig 2 and table 2 above, and TRMo-2 is about 50,000x less active. Consequently TRMo-2 has a specific activity of about 0.006 units per mg in terms of stimulation of cyclic AMP production. Again, this is well outside any of the numerical limits of the claims.

Thus, while the Yoshida antibody TRMo-2 may have worked to some de minimus extent, even if it could be reproduced, it comes nowhere close to meeting the numerical limitations of the present claims.

Accordingly, Applicants submit that the rejections under 35 USC § 102 are in error and should be withdrawn.

Obviousness rejection

Claims 204-209, 212 and 213 stand rejected as obvious over Yoshida, in view of UniProt P16473 disclosing the amino acid sequence of the human TSH Receptor, Zhong and Kohn as evidenced by WO 91/09137. The secondary references relate to the techniques for determining sequences of monoclonal antibodies and thus to make recombinant antibodies, and to motivation to make the recombinant antibody as now claimed. These techniques are useless, however without possession by the public of the Yoshida antibody to perform these steps on, Thus, the lack of enablement is fatal with respect to this assertion of obviousness. It is not possible to determine from anything that Yoshida disclosed or enabled the sequence of the Yoshida antibody TRMo-2, and the evidence of motivation in this case is more properly understood as a long felt need that neither Yoshida nor anyone else in the art had answered prior to the present invention. Furthermore, the comments above with respect to the activity levels are relevant to this rejection as well. There is simply no basis to imagine that a recombinant antibody based on the TRMo-2 would have a level of activity anywhere near the levels required by the claims.

Thus, the rejection under 35 USC § 103(a) should be withdrawn.

Written Description

Claims 134, 135, 200-202, 210 and 211 which recite specific sequences are rejected as lacking written description. The basis for this rejection is that the claims recite antibodies in which parts of the structure are identified as having specific sequences while other parts remain generic. The Examiner asserts that the specification "does not provide an adequate description of the structure of the antibodies such that the skilled artisan would be aware was in possession of the genera of claimed antibodies." Applicants respectfully disagree.

As a first matter, it should be noted that the claims are directed to antibodies or **fragments**. Each of the sequences is the sequence of an antibody fragment, and Applicant clearly had possession of these fragments independent of any association with a complete antibody structure as reflected in the sequence listing and original claims 22- 27 of the PCT application as filed.. Applicants further point out that the application, starting on Page 9 and continuing to Page 10, states that the binding partner can have a V_H domain without a V_L domain, or vice versa. Thus, the application expressly discloses molecules that comprise just one type of domain, as well as both types of domain. Furthermore, the application states (starting on Page 11, that one or more CDRs can be incorporated into a suitable framework, such as a different antibody, to confer the binding and stimulating properties of the recited antibody.

It is further noted it is well known that the constant domains of human heavy chain and the lambda light chain constant regions are both essentially constant and also known. See Kabat

et al Sequences of proteins of immunological interest (US Public Health Service, Bethesda, MD) (1991) Vol 1. Copies of some relevant pages are attached as Ex. A. In addition, human CH1 sequence data is available from Genbank (accession number A49444). In this case, the amino acid sequence of hMAb TSHR1 heavy chain amino acids 1-131 are shown (SEQ ID No 5). Amino acids 1-121 are the variable region (VH) domain (SEQ ID No 1) and amino acids 122-131 are part of the constant region (CH1). Similarly, light chain amino acids 1-111 are the variable region domain of the light chain (VL).

Furthermore, the purpose of the written description requirement is to ensure that the inventor had possession of the invention, not to require, or even encourage a restatement of all the background in the relevant art. In this case, persons skilled in the art are well aware of the basic structure of antibodies. For example, the Zhong abstract cited by the Examiner shows taking a light chain V_L from a monoclonal antibody, sequencing it, and says it can be used with other unspecified elements to make a recombinant antibody. Since the Examiner concluded that this was sufficient to show a person skilled in the art that making recombinant antibodies from monoclonal antibodies was routine, the assertion that more than the important sequences must be disclosed by a patent application is unfounded.

In this regard, the examiner is directed to the decision of the Federal Circuit in *Capon v. Eshhar*, 76 USPQ2d 1078 (Fed. Cir. 2005). In *Capon*, the Federal Circuit has considered an interference proceeding, in which the Patent Office Board of Appeals found that neither applicants' disclosure met the written description requirement Both applications related to chimeric genes designed to combine DNA encoding known antigen-binding domains and known lymphocyte-receptor protein into a unitary gene. Both applications claims such chimeric genes generically. The Patent Office Board of Appeals and Interferences held that there was a lack of written description because the applications claimed the invention in terms of function, instead of specific sequences or structures. The Federal Circuit vacated this holding, finding that the failure to explicitly present specific sequences based on known genes did not create a basis for a rejection for lack of written description, and stated that no *per se* rule exists for the recitation of specific sequences.

Here, the structure of complete antibodies is well known, but it is also known that the various fragments (V_H V_L and CDR's) recited in the claims are what give rise to specificity for an antigen in this case TSH receptor. This is the meat that gives the binding specificity and the examiner has offered no reason (other than form paragraphs) why a person skilled in the art would not see a disclosure of the invention as claimed in the application as filed. Accordingly, there is no basis for the written description rejection as presented, and it should be withdrawn.

Notwithstanding this view, Applicants enclose several complete articles from the art relating generally to the structure and construction of antibodies to show that this is indeed

known to persons skilled in the art, and request that these be made of record and taken into account by the Examiner and that this knowledge be fully addressed should the rejection be repeated. Liang et al. Journal of Immunological Methods 2001 247: 119-130 (Exhibit C) shows that in the same vector different human heavy chain V region genes can be assembled with the same CH1, CH2 and CH3 domain genes and an appropriate lambda light chain V region gene with the same Cl gene. These recombinant antibodies can be expressed and their antigen binding function is preserved. A similar approach was also described in Knappik et al. Journal of Molecular Biology (2000) 296: 57-86 (Exhibit D).

Obviousness-type Double Patenting

Applicants submit that the USPTO lacks any authority to make an obviousness-type double patenting rejection in the absence of a controlling statute and properly promulgated rules. However, since the cited application is not yet acted upon and this case should now be allowed, this issue is moot at this time.

Conclusion

For the foregoing reasons, the considered claims of this application are believed to be in form for allowance. Recombination of the non-elected claims, as appropriate, is requested. To the extent claims will not be recombined, a telephone call to the undersigned will result in authorization to cancel non-elected claims.

Respectfully submitted,

Marina T. Larson Ph.D.
PTO Reg. No. 32,038
Attorney for Applicant
(970) 262-1800

NIBSC
Assuring the quality of biological medicines

Blanche Lane
South Mimms
Potters Bar
Hertfordshire EN6 3QG
United Kingdom

THYROID STIMULATING ANTIBODY 1ST INTERNATIONAL STANDARD 90/672, 0.1 INTERNATIONAL UNITS/AMPOULE Instructions for use

1. INTRODUCTION

The International Standard for thyroid stimulating antibody (TSAb) consists of a batch of ampoules containing freeze-dried plasma proteins from a single human patient with high TSAb levels. The preparation has been evaluated in an international collaborative study and shown to possess both thyroid stimulating and thyroid receptor binding activity. At its 46th meeting in 1995 the Expert Committee on Biological Standardization of WHO established the preparation coded 90/672 as the International Standard for thyroid stimulating antibody, but noted that whilst the preparation is suitable for both bioassays and receptor assays of anti-thyroid receptor autoantibodies, the unitage in terms of 90/672 may not be equivalent to earlier standards such as the MRC LATS-B standard 65/122.

2. AMPOULE CONTENTS

Each ampoule contains the freeze-dried residue of 1.0ml of a solution which contained:

0.02M phosphate buffer
dialysed human plasma proteins

3. UNITAGE

0.1 International Units (100 milli-International Units) per ampoule by definition.

4. CAUTION

4.1 THIS PREPARATION IS NOT FOR ADMINISTRATION TO HUMANS.

4.2 A safety data sheet is included in the last page of these instructions

4.3 The preparation contains material of human origin which has been tested and found negative for HB_sAg and anti-HIV. However, as with all preparations of human origin, this material cannot be assumed to be free from infectious agents. Suitable precautions should be taken in the use and disposal of the ampoule and its contents.

5. USE OF AMPOULES

For all practical purposes each ampoule contains the same amount of the above materials. Dissolve the total contents in a known amount of suitable buffer solution with carrier protein (free of peptidase), where extensive dilution is required, to minimise loss by surface adsorption.

No attempt should be made to weigh out portions of the freeze-dried powder.

For economy of use it is recommended that the solution be subdivided into several small containers, which are frozen rapidly to below -70°C and then stored below -30°C in the dark. Repeated freezing and thawing should be avoided.

The material has not been sterilized and contains no bacteriostat.

Unopened ampoules should be stored below -20°C.

6. DIRECTIONS FOR OPENING AMPOULE

a) Tap ampoule gently to collect the material at the bottom end.

b) Score the ampoule all the way round the circumference near the top using a sharp ampoule file. Heat a thin glass rod to white heat and apply firmly at the hot end to the file score. If a crack does not appear, deepen the file score. Reheat the glass rod and re-apply. When a crack appears, hold the ampoule almost horizontally and gently remove the top (empty) portion.

c) Take care that no particles of glass fall into the ampoule and no material is lost from the ampoule.

7. PREPARATION OF AMPOULES

7.1 Bulk material

The preparations of TSAb-rich plasma was obtained from a pregnant patient exhibiting high levels of TSAb, whose plasma was regularly exchanged during pregnancy, and was kindly provided by Professor D.S. Munro, Northern General Hospital, Sheffield. The material, containing citrate and traces of heparin, was donated to NIBSC in 1975, and has been stored at -70°C since that time. The preparation was tested at NIBSC for HIV antibodies, and Hepatitis-B antigen, and was negative.

7.2 Preparation of ampoules

The preparation was dialysed against 0.02M sodium phosphate, clarified by centrifugation, and filtered through a 0.45μ filter before being dispensed (1ml aliquots) into glass ampoules, and lyophilized and secondary desiccated according to the procedures described by WHO for International Biological Standards (Annex 4, 29th ECBS report, WHO Tech. Rep. Ser. No 626, 1978).

8. COLLABORATIVE STUDY

8.1 Design of the study

Nine laboratories in five countries participated in the study. Participants were requested:

1 To examine by bioassay and by receptor binding assay two candidate preparations of TSAb-rich plasma, and to compare them with standards currently in use by the participants.

2 To calibrate each of the two preparations in terms of MRC LATS-B, and in terms of local standards, and to assign a unitage to each.

3 To assess the stability of each of the two preparations by comparing them with ampoules that had been subjected to accelerated thermal degradation.

8.2 Results and conclusions

The preparation coded 90/672 exhibited activity in both in vitro bioassays for thyroid stimulating antibody activity, and in thyroid binding receptor assays. In terms of the MRC research standard B for long-acting thyroid stimulator, 65/122, estimates by bioassay were in the range 10-40 mU/ampoule, and by receptor assay were in the range 250-450 mU/ampoule. The stability of the preparation was acceptable.

The preparation was formally established as the International Standard for thyroid stimulating antibody at the 1995 meeting of the Expert Committee on Biological Standardization of WHO, and assigned a content of 100 milli-Units per ampoule.

In using the preparation, with its provisionally assigned unitage, users should note:

- The ampoule content does not represent a formal continuity of unitage with 65/122 or any other standard for TSAb. The unitage is only approximately equivalent to that of 65/122, and the relationship between units of 65/122 and 90/670 will vary widely depending on the assay system employed.

- The preparation is derived from a single patient, and may not be qualitatively suitable to serve as a standard for all TSAb samples.

9. PRODUCT LIABILITY

9.1 Information emanating from NIBSC is given after the exercise of all reasonable care and skill in its compilation, preparation and issue, but is provided without liability in its application and use.

9.2 This product is intended as for use as a standard or reference material in laboratory work in relation to biological research, manufacturing or quality control testing of biological products, or in the field of in vitro diagnostics. It is the

responsibility of the user to ensure that he/she has the necessary technical skills to determine the appropriateness of the product for the proposed application. Results obtained from this product are likely to be dependent on the conditions of use and the variability of materials beyond the control of NIBSC.

NIBSC accepts no liability whatsoever for any loss or damage arising from the use of this product, whether loss of profits, or indirect or consequential loss or otherwise, including, but not limited to, personal injury other than as caused by the negligence of NIBSC. In particular, NIBSC accepts no liability whatsoever for :-
i) results obtained from this product; and/or
ii)non-delivery of goods or for damages in transit.

9.3 In the event of any replacement of goods following loss or damage a customer accepts as a condition of receipt of a replacement product, acceptance of the fact that the replacement is not to be construed as an admission of liability an NIBSC's behalf.

10. ACKNOWLEDGEMENTS

Acknowledgements are due to Professor Donald Munro for providing the plasma sample.

11. CITATION

To avoid scientific confusion in publications or data sheets for assay kits in which this Standard has been used for primary calibration, the Standard should be cited by its correct title and ampoule code. The correct address of this Institute which distributes the Standard should also be given.

12. REFERENCES

WHO Expert Committee on Biological Standardization (1989). Guidelines for the preparation, characterization and establishment of international and other standards and reference reagents for biological substances. WHO Technical Report Series 800, 181-213 (1990).

INTERNATIONAL STANDARD FOR THYROID STIMULATING ANTIBODY
NIBSC Code: 90/672
Instructions for Use (Version November 20, 1995 Initial Version)

MATERIAL SAFETY SHEET

Physical Properties (at room temperature)	
Physical appearance	*Solid*
Fire hazard	*None*

Chemical Properties			
Stable	*Yes*	Corrosive	*No*
Hygroscopic	*Yes*	Oxidising	*No*
Flammable	*No*	Irritant	*No*
Other (specify)	*Contains material of human origin*		
Handling:	*See precautions in section 4.3*		

Toxicological Properties	
Effects of inhalation	*No adverse effects have been reported for this material*
Effects of ingestion	*No adverse effects have been reported for this material*
Effects of skin absorption	*No adverse effects have been reported for this material*

Suggested First Aid	
Inhalation	*Seek medical advice*
Ingestion	*Seek medical advice*
Contact with eyes	*Wash with copious amounts of water. Seek medical advice.*
Contact with skin	*Wash thoroughly with water*

Action on Spillage and Method of Disposal
Spillages of vial contents should be taken up with absorbent material wetted with a viricidal agent. Rinse area with a viricidal agent followed by water. *Absorbent material used to treat spillages should be treated as biologically hazardous waste.*

Compiled by:	Date: 21/12/95

From: Kabat et al. Sequences of proteins of immunological interest (US Public Health Service, Bethesda, MD) (1991) Vol 1

Page 718	Heavy chain description of sequences shown on page 662.
Page 662	Sequences 28-40 are human IgG 1 CH-1 domains.
Page 658	Light chain description of sequences shown on pages 653 and 654.
Pages 653 & 654	Sequences 1-28 are human lambda constant chains.

Ex B

ALLOTYPE: HEAVY CONSTANT CHAINS

78) B1-8.DELTA1: IGD-H (C57BL/6 ORIGIN)
83) MOPC21: IG1A
86) IGG2B(A)'CL': A
87) IGG2B(A)'CL': A
88) IGG2B(B)'CL': B
90) MPC11: IG1A
92) IGG2A(A)'CL': A
94) IGG2A(B)'CL': B
97) MOPC173: IG1A
98) CBPC191: IG1B
130) CT-12: E15(I+J+)
131) PTKA: E15(I+J-)
132) HA-3: E15(I+J+)
133) HA-11: E15(I+J+)
134) HA-1: E15(I+J-)
135) HA-8: E15(I+J-)
136) HA-42: E15(I-J-)

CLASSIFICATION: HEAVY CONSTANT CHAINS

1) HUMAN IGM'CL': HUMAN IGM
2) HUMAN IGM' 'CL: HUMAN IGM
3) GAL: HUMAN IGM
4) OU: HUMAN IGM
5) NOT: HUMAN IGM
7) GLI: HUMAN IGM
8) HUMAN IGM MEMB'CL: HUMAN IGM
9) WAH: HUMAN IGD
10) NIG-65: HUMAN IGD
11) ERI: HUMAN IGD
14) HUMAN IGG3'CL': HUMAN IGG3
15) OMM'CL: HUMAN IGG3
16) HER: HUMAN IGG3
17) PRO: HUMAN IGG3
18) JON: HUMAN IGG3
19) WIS: HUMAN IGG3
20) SPA: HUMAN IGG3
21) ZUC: HUMAN IGG3
22) ZUC': HUMAN IGG3
23) ZIP: HUMAN IGG3
24) BAD: HUMAN IGG3
25) JIR: HUMAN IGG3
26) CHA: HUMAN IGG3-1G31
27) GOK: HUMAN 1GG3-IG31
28) ZU: HUMAN IGG1
29) MIL: HUMAN IGG1
30) CHA: HUMAN IGG1
31) VAJ: HUMAN IGG1
32) LEB: HUMAN IGG1
33) REI: HUMAN IGG1
34) TOK: HUMAN IGG1
35) ZAC: HUMAN IGG1
37) KOL: HUMAN IGG1
38) MCG: HUMAN 1GG1
39) LEC: HUMAN IGG1
40) DOB: HUMAN IGG1
41) ROH: HUMAN IGA1
42) TRO: HUMAN IGA1
43) CAR: HUMAN IGA1
47) HUMAN IGG'CL: HUMAN IGG
48) HUMAN IGG2'CL: HUMAN IGG2
49) YEL: HUMAN IGG2
50) KIE: HUMAN IGG2
51) SA: HUMAN IGG2
52) PIG: HUMAN IGG2
53) HUMAN IGG4'CL: HUMAN IGG4
54) VIN: HUMAN IGG4
55) HUMAN IGE'CL: HUMAN IGE
57) CHIMP IGE'CL: CHIMPANZEE IGE
58) ORANGUTAN IGE'CL: ORANGUTAN IGE
59) UZ66'CL: HUMAN IGE
60) MD: HUMAN IGE
61) NOT: HUMAN IGA2 A2M(2)
62) LAN: HUMAN IGA2 A2M(1)
63) HUMAN IGA2'CL: HUMAN IGA2M
64) IGM'CL: MOUSE IGM
67) IGM'CL': MOUSE IGM
68) IGM-b'CL: C57BL/6 MOUSE IGM
69) MUTANT 102'CL: MOUSE IGH
70) MOPC104E: MOUSE IGM
71) MOPC104E MEMB'CL: MOUSE IGM MEMBRANE BOUND
72) HPC76'CL: MOUSE IGM
73) IGD'CL: MOUSE IGD
76) IGD SECR'CL: MOUSE IGD SECRETED EXON
77) IGD MEMB'CL: MOUSE IGD MEMBRANE EXON
78) B1-8.DELTA1: MOUSE IGD
79) MOUSE IGG3'CL: MOUSE IGG3
81) IGG1'CL: MOUSE IGG1
82) IGG1 MEMB'CL: MOUSE IGG1 MEMBRANE BOUND
83) MOPC21: MOUSE IGG1
84) IF2'CL: MOUSE IGG1 WITH Ch1 DELETED
85) IGH 11.19.3: MOUSE IGG2B/IGG2A HYBRID
86) IGG2B(A)'CL: MOUSE (BALB/C) IGG2B
87) IGG2B(A)'CL': MOUSE (BALB/C) IGG2B
88) IGG2B(B)'CL: MOUSE (C57BL/6) IGG2B
89) IGG2B MEMB'CL: MOUSE IGG2B MEMBRANE BOUND
90) MPC11: MOUSE IGG2B
92) IGG2A(A)'CL: MOUSE (BALB/C) IGG2A
93) 17/9'CL: MOUSE IGG2A
94) IGG2A(B)'CL: MOUSE (C57BL/6) IGG2A
95) IGG2A(B)'CL': MOUSE (C57BL/6) IGG2A
96) IGG2A MEMB'CL: MOUSE IGG2A MEMBRANE BOUND

HEAVY CONSTANT CHAINS CH1 REGION (cont'd)

	EU INDEX	OU INDEX	20 SPA #	21 ZUC #	22 ZUC' #	23 KUP #	24 BRU #	25 JIR	26 CHA #	27 GOE	28 EU	29 NIE #	30 CRA #	31 VAU #	32 LEB #	33 EST #	34 YOK #	35 SAC	36 HUMAN IGG1 'CL #	37 KOL	38 MCG	39 LEC	40 DOB	41 BUR #	42 TRO	43 CAR #
113A			---	---	---				---		---	---		---	---	---	---	---	---	---				---	---	
113B			---	---	---				---		---	---		---	---	---	---	---	---	---				---	---	
113C			---	---	---				---		---	---		---	---	---	---	---	---	---				---	---	
113D			---	---	---				---		---	---		---	---	---	---	GLX	---	---				---	---	
113E			---	---	---				---		---	---		---	---	---	---	SER	---	---				---	---	
114	118	127	---	---	---				---		ALA	ALA	---	---	---	---	---	ALA	ALA	ALA				ALA	ALA	
115	119	128	---	---	---				---		SER	SER	---	---	---	---	---	SER	SER	SER				SER	SER	
116	120	129	---	---	---				---		THR	THR	---	---	---	---	---	THR	THR	THR				PRO	PRO	
117	121	130	---	---	---				---		LYS	LYS	---	---	---	---	---	LYS	LYS	LYS				THR	THR	
118	122	131	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				SER	SER	
119	123	132	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				PRO	PRO	
			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
120	124	133	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				LYS	LYS	
121	125	134	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				VAL	VAL	
122	126	135	---	---	---				---		PHE	PHE	---	---	---	---	---		PHE	PHE				PHE	PHE	
123	127	136	---	---	---	PRO	PRO		---		PRO	PRO	---	---	---	---	---		PRO	PRO				PRO	PRO	
124	128	137	---	---	---	LEU	LEU		---		LEU	LEU	---	---	---	---	---		LEU	LEU				LEU	LEU	
125	129	138	---	---	---	ALA	ALA		---		ALA	ALA	---	---	---	---	---		ALA	ALA				SER	SER	
126	130	139	---	---	---	PRO	PRO		---		PRO	PRO	---	---	---	---	---		PRO	PRO				LEU	LEU	
127	131	140	---	---	---	CYS	CYS		---		SER	SER	---	---	---	---	---		SER	SER				CYS	CYS	
128	132	141	---	---	---	SER	SER		---		SER	SER	---	---	---	---	---		SER	SER				SER	SER	
129	133	142	---	---	---	ARG	ARG		---		LYS	LYS	---	---	---	---	---		LYS	LYS				THR	THR	
130	134	143	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				GLX	GLN	
131			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
132			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
133	135	144	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				---	---	
134	136	145	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				PRO	PRO	
135	137	146	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				ASX	ASP	
136	138	147	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				GLY	GLY	
137	139	148	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				ASX	ASN	
138	140	149	---	---	---				---		ALA	ALA	---	---	---	---	---		ALA	ALA				VAL	VAL	
139	141	150	---	---	---				---		ALA	ALA	---	---	---	---	---		ALA	ALA				VAL	VAL	
140	142	151	---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				ILE	ILE	
141	143	152	---	---	---	GLY	GLY		---		GLY	GLY	---	---	---	---	---		GLY	GLY				ALA	ALA	
142	144	153	---	---	---	CYS	CYS		---		CYS	CYS	---	---	---	---	---		CYS	CYS				CYS	CYS	
143	145	154	---	---	---	LEU	LEU		---		LEU	LEU	---	---	---	---	---		LEU	LEU				LEU	LEU	
144	146	155	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				VAL	VAL	
145	147	156	---	---	---				---		LYS	LYS	---	---	---	---	---		LYS	LYS				GLN	GLN	
146	148	157	---	---	---				---		ASP	ASP	---	---	---	---	---		ASP	ASP				GLY	GLY	
147	149	158	---	---	---				---		TYR	TYR	---	---	---	---	---		TYR	TYR				PHE	PHE	
148	150	159	---	---	---				---		PHE	PHE	---	---	---	---	---		PHE	PHE				PHE	PHE	
149	151	160	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				PRO	PRO	
150	152	161	---	---	---				---		GLU	GLN	---	---	---	---	---		GLU	GLU				GLN	GLN	
151	153	162	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				GLN	GLN	
152	154	163	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				PRO	PRO	
153	155	164	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				LEU	LEU	
154	156	165	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				SER	SER	
155			---	---	---				---		---	---	---	---	---	---	---		---	---				VAL	VAL	
			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
156	157	166	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				THR	THR	
157	158	167	---	---	---				---		TRP	TRP	---	---	---	---	---		TRP	TRP				TRP	TRP	
158		168	---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
159		169	---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
160		170	---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
161		171	---	---	---				---		---	---	---	---	---	---	---		---	---				SER	SER	
162	159	172	---	---	---				---		ASN	ASN	---	---	---	---	---		ASN	ASN				GLU	GLU	
163	160	173	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				SER	SER	
164	161	174	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				GLY	GLY	
165	162	175	---	---	---				---		ALA	ALA	---	---	---	---	---		ALA	ALA				GLX	GLN	
166	163		---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				---	---	
167	164		---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				---	---	
168	165	176	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				GLY	GLY	
169	166	177	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				VAL	VAL	
170			---	---	---				---		---	---	---	---	---	---	---		---	---				THR	THR	
171	167	178	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				ALA	ALA	
172	168	179	---	---	---				---		HIS	HIS	---	---	---	---	---		HIS	HIS				ARG	ARG	
173	169	180	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				ASX	ASN	
174	170	181	---	---	---				---		PHE	PHE	---	---	---	---	---		PHE	PHE				PHE	PHE	
175	171	182	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				PRO	PRO	
176	172	183	---	---	---				---		ALA	ALA	---	---	---	---	---		ALA	ALA				PRO	PRO	
177	173	184	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				SER	SER	
178	174	185	---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				GLX	GLN	
179	175	186	---	---	---				---		GLN	GLN	---	---	---	---	---		GLN	GLN				ASX	ASN	
180	176	187	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				ALA	ALA	
181			---	---	---				---		---	---	---	---	---	---	---		---	---				SER	SER	
182	177		---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				GLY	GLY	
			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
			---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
183	178	188	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				ASX	ASN	
184	179	189	---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				LEU	LEU	
185	180	190	---	---	---				---		TYR	TYR	---	---	---	---	---		TYR	TYR				TYR	TYR	
186	181	191	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				THR	THR	
187	182	192	---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				THR	THR	
188	183	193	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				SER	SER	
189	184	194	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				SER	SER	
190	185	195	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				GLN	GLN	
191	186	196	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				LEU	LEU	
192	187	197	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				THR	THR	
193	188	198	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				LEU	LEU	
194	189	199	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				PRO	PRO	
195	190	200	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				ALA	ALA	
196	191	201	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				THR	THR	
197	192	202	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				GLX	GLN	
198	193	203	---	---	---				---		LEU	LEU	---	---	---	---	---		LEU	LEU				CYS	CYS	
199	194	204	---	---	---				---		GLY	GLY	---	---	---	---	---		GLY	GLY				LEU	LEU	
200	195	205	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				ALA	ALA	
201		206	---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
202		207	---	---	---				---		---	---	---	---	---	---	---		---	---				---	---	
203	196	208	---	---	---				---		GLN	GLN	---	---	---	---	---		GLN	GLN				GLY	GLY	
204		209	---	---	---				---		---	---	---	---	---	---	---		---	---				LYS	LYS	
205	197	210	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				SER	SER	
206	198	211	---	---	---				---		TYR	TYR	---	---	---	---	---		TYR	TYR				VAL	VAL	
207	199		---	---	---				---		ILE	ILE	---	---	---	---	---		ILE	ILE				THR	THR	
208	200	212	---	---	---				---		CYS	CYS	---	---	---	---	---		CYS	CYS				CYS	CYS	
209	201	213	---	---	---				---		ASN	ASN	---	---	---	---	---		ASN	ASN				HIS	HIS	
210	202	214	---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL				VAL	VAL	
211	203	215	---	---	---				---		ASN	ASN	---	---	---	---	---		ASN	ASN				LYS	LYS	
212	204	216	---	---	---				---		HIS	HIS	---	---	---	---	---		HIS	HIS				HIS	HIS	
213	205	217	---	---	---				---		LYS	LYS	---	---	---	---	---		LYS	LYS				TYR	TYR	
214	206	218	---	---	---				---		PRO	PRO	---	---	---	---	---		PRO	PRO				THR	THR	
215	207	219	---	---	---				---		SER	SER	---	---	---	---	---		SER	SER				ASN	ASN	
216	208	220	---	---	---				---		ASN	ASN	---	---	---	---	---		ASN	ASN				PRO	PRO	
217	209	221	---	---	---				---		THR	THR	---	---	---	---	---		THR	THR				SER	SER	
218	210	222	---	---	---				---		LYS	LYS	---	---	---	---	---		LYS	LYS	LYS	LYS		GLN	GLN	
219	211		---	---	---				---		VAL	VAL	---	---	---	---	---		VAL	VAL	VAL		VAL	---	---	
219A			---	---	---				---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	
220	212	223	---	---	---				---		ASP	ASP	---	---	---	---	---		ASP	ASP	ASP		ASP	ASX	ASP	
221	213	224	---	---	---				---		LYS	LYS	---	---	---	---	---		LYS	LYS	LYS	LYS	LYS	VAL	VAL	
222	214	225	---	---	---				---		ARG	LYS	---	---	---	---	---		LYS	ARG	ARG	ARG	LYS	THR	THR	
223	215	226	---	---	---				---		---	VAL	---	---	---	---	---		---	---	VAL	VAL	VAL	VAL	VAL	
223A		227	---	---	---				---		---	---	---	---	---	---	---		---	---	---	---	---	PRO	PRO	
223B		228	---	---	---				---		---	---	---	---	---	---	---		---	---	---	---	---	CYS	CYS	
223C		229	---	---	---				---		---	---	---	---	---	---	---		---	---	---	---	---	PRO	PRO	

CLASSIFICATION: LAMBDA LIGHT CONSTANT CHAINS

1) NEWM: HUMAN LAMBDA
2) NEW: HUMAN LAMBDA
3) WAH: HUMAN LAMBDA
4) MOH: HUMAN LAMBDA
5) NIG-84: HUMAN LAMBDA
6) KU492: HUMAN LAMBDA
7) SH: HUMAN LAMBDA
8) VIL: HUMAN LAMBDA
9) LEVLAMBDA'CL: HUMAN LAMBDA
10) HIL: HUMAN LAMBDA
11) X: HUMAN LAMBDA
12) KERN: HUMAN LAMBDA
13) NIG-48: HUMAN LAMBDA
14) MOT: HUMAN LAMBDA
15) LAMBDA 6'CL: HUMAN LAMBDA
16) SH: HUMAN LAMBDA
17) NG-3C4'CL: HUMAN LAMBDA SUSGRQUE IV
18) BO: HUMAN LAMBDA
19) MCG: HUMAN LAMBDA
20) AMYLOID-AR: HUMAN LAMBDA FRAGMENT
21) SUT: HUMAN LAMBDA
22) TRO: HUMAN LAMBDA
23) MOR: HUMAN LAMBDA
24) KK-OS-'CL: HUMAN LAMBDA
25) KK-OS+'CL: HUMAN LAMBDA
26) MCG'CL: HUMAN LAMBDA
27) 14.1'CL: HUMAN LAMBDA
28) 16.1'CL: HUMAN LAMBDA
29) MOUSE 12'CL: MOUSE LAMBDA
30) MOUSE CL2S'CL: MOUSE (STRAIN SPE.SPE) LAMBDA
31) 1C4D6.5'CL: MOUSE (STRAIN CLA-2/Cn) LAMBDA
32) MOUSE 14'CL: MOUSE LAMBDA
33) MOUSE 21'CL: MOUSE LAMBDA
34) MOUSE 23'CL: MOUSE LAMBDA
35) MOPC104E: MOUSE LAMBDA 1
36) FLAT-13'CL: MOUSE LAMBDA
37) IS 3031AMBDA'CL: MOUSE LAMBDA
38) S43'CL: MOUSE LAMBDA 1
39) MOPC315: MOUSE LAMBDA 2
40) KL.82A83.12'CL: MOUSE LAMBDA 1
41) MOPC315A'CL: MOUSE LAMBDA
42) 6-2: MOUSE LAMBDA 3 (WITH VARIABLE REGION BELONG TO LAMBDA II)
43) MOPC315-26'CL': MOUSE LAMBDA
44) CBPC-49: MOUSE LAMBDA 3
45) S-8: MOUSE LAMBDA 3
46) T31'CL: MOUSE LAMBDA LIGHT CHAIN
47) IG 2SLAMBDA'CL: MOUSE LAMBDA
48) BJL/G'CL: MOUSE LAMBDA 1 OF STRAIN BJL/J
49) pOR7'CL: BASILEA RABBIT LAMBDA
50) pOR8'CL: BASILEA RABBIT LAMBDA
51) 4560: BASILEA RABBIT LAMBDA
52) BASILEA RABBIT LAMBDA PCL14'CL: BASILEA RABBIT LAMBDA
53) 4431(RAB): RABBIT LAMBDA
54) RABBIT LAMBDA 1'CL: RABBIT LAMBDA
55) RABBIT LAMBDA 2'CL: RABBIT LAMBDA
56) RABBIT LAMBDA 3'CL: RABBIT LAMBDA
57) RABBIT LAMBDA 4'CL: RABBIT LAMBDA
58) RB-1: RABBIT LAMBDA
59) LAMBDA POOL: RABBIT LAMBDA
60) GPLA: GUINEA PIG LAMBDA
61) GPLA1: GUINEA PIG LAMBDA-1
62) GPLA2: GUINEA PIG LAMBDA-2
63) GPLA3: GUINEA PIG LAMBDA-3
64) PIGLA: PIG LAMBDA
65) HORLA: HORSE LAMBDA
66) GOATLA: GOAT LAMBDA
71) BOVLA: BOVINE LAMBDA
72) BUFLA: WATER-BUFFALO LAMBDA
73) DOGLA: DOG LAMBDA
75) S11'CL: CHICKEN LAMBDA
76) 243'CL: CHICKEN LAMBDA
77) S18'CL: CHICKEN LAMBDA
78) S13'CL: CHICKEN LAMBDA
79) S11'CL: CHICKEN LAMBDA
80) S7'CL: CHICKEN LAMBDA
81) S5'CL: CHICKEN LAMBDA
82) CHILA: CHICKEN LAMBDA
83) TURLA: TURKEY LAMBDA

REFERENCE: LAMBDA LIGHT CONSTANT CHAINS

1) NEWM: CHEN,H.L. & POLJAK,R.J. (1974) BIOCHEMISTRY,13,1295-1302. (CHECKED BY AUTHOR 07/16/79)
2) NEW: LANGER,R.,STEINMETZ-KAYNE,M. & HILSCHMANN,N. (1968) Z.PHYSIOL.CHEM.,349,945-951.
3) WAH: TAKAHASHI,N.,TAKAHASHI,M.,TETAERT,D. & PUTNAM,F.W. (1985) PROC.NAT.ACAD.SCI.USA,80,3688-3690. (CHECKED BY AUTHOR 06/15/85)
4) MOH: MIHAESCO,E.,ROY,J.-P.,CONGY,N.,PERAN-RIVAT,L. & MIHAESCO,C. (1985) ESR.J.BIOCHEM.,150,349-357.
5) NIG-84: TONOUE,H.,KAMETANI,F.,NOSHI,A.,SHINODA,T. & ISOBE,T. (1985) FEBS LETTERS,185,139-141.
6) KU492: CVETENAR,T.,STEITZEN,K. & NEUTZMANN,P. (1985) BIOCHEM.J.,232,183-190.
7) SH: CARVER,F.J.,CHANG,L.,MENDICINO,J.,ISOBE,T. & OSSERMAN,E.F. (1975) PROC.NAT.ACAD.SCI.USA,72,4559-4563. (CHECKED BY AUTHOR 05/31/83)
8) VIL: PODSTINGL,H. & HILSCHMANN,N. (1969) Z.PHYSIOL.CHEM.,350,1148-1152; (1971) Z.PHYSIOL.CHEM.,352,869-877.
9) LEVLAMBDA'CL: ANDERSON,M.L.,BROWN,L.,MCKENZIE,K.,KELLOW,J.E. & YOUNG,B.D. (1985) NUC.ACIDS RES.,13,6,2931-2941. (CHECKED BY AUTHOR 08/21/85)
10) HIL: LOPES DE CASTRO,J.A., CHIU,Y.Y.H. & POLJAK,R.J. (1978) BIOCHEMISTRY,17,1718-1723. (CHECKED BY AUTHOR 07/16/79)
11) X: HILETZIN,C. CLEOL,J.B. & JARVIS,J.M. (1968) BIOCHEM.J.,110,631-652.
12) KERN: PORSTINGL,H.,HESS,M. & HILSCHMANN,N. (1971) Z.PHYSIOL.CHEM.,352,247-256.
13) NIG-48: SHINODA,T.,YOSHIMURA,K. & KAMETANI,F. (1983) BIOCHEM.BIOPHYS.RES.COMMUN.,117,587-592.
14) MOT: ROJIDA,H.,OOARI,S. & INDEHARA,T. (1980) MOL.IMMUNOL.,17,1467-1474.
15) LAMBDA 6'CL: DARIAVACH,P.,LEFRANC,G. & LEFRANC,M.-P. (1987) PROC.NATL.ACAD.SCI.USA,84,9074-9078.
16) SH: TITANI,K.,WIKLER,M.,SHINODA,T. & PUTNAM,F.W. (1970) J.BIOL.CHEM.,245,2171-2176. (CHECKED BY AUTHOR 06/15/83)
17) NG-3C4'CL: YAMASAKI,N.,KONDO,S. & MIYAMARE,Y. (1983) MOL.IMMUNOL.,20,981-985.
18) BO: WIKLER,M. & PUTNAM,F.W. (1970) J.BIOL.CHEM.,245,4498-4507. (CHECKED BY AUTHOR 06/15/83)
19) MCG: FETT,J.W. & BENTSCH,H.F. (1974) BIOCHEMISTRY,13,4102-4114. (CHECKED BY AUTHOR 07/20/79)

LAMBDA LIGHT CONSTANT CHAINS

	INVARIANT RESIDUES	1 NEWM	2 NEW #	3 WAH #	4 MEM #	5 NIG -84	6 ES492	7 SM	8 VIL #	9 LBV LAMBDA 'CL	10 HIL	11 X #	12 KERN #	13 NIG -68	14 MOT	15 LAMBDA 6 'CL #	16 SH #	17 H6-3C4 'CL	18 BO	19 MCG	20 AMYLOID -AR	21 SUT	22 THO	23 HOR
107A		---	---	---	---	---	---	---	---	---	---	---	---	GLY	---	GLY	---	---	---	---	---	---	---	---
108	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	---	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN
109	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	---	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
109A		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
110		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
111		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
112		ALA	ALA	ALA	ASN	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ASN	ALA	ALA	ALA	ALA	ASN	ALA	ALA	ALA	ALA
113	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
114		SER	SER	SER	THR	SER	SER	SER	SER	SER	SER	SER	SER	SER	THR	SER	SER	SER	SER	THR	SER	SER	SER	SER
115		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL
116		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR
117		LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
118	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE
119	PRO(.99)	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
120	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
121	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
122		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
123	GLU(.97)	GLX	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLX	GLU	GLU	GLU	GLU	GLU
124	GLU(.97)	GLX	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLX	GLU	GLU	GLU	GLU	GLU
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
124A	LEU	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
125		LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
126		GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLX	GLN	GLN	GLN	GLN	GLN
127		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
128		ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASX	ASN	ASN	ASN	ASN	ASN
129	LYS(.99)	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	ARG	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
130	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
131	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR
132	LEU(.99)	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
133	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL
134	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
135		LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
136	ILE(.96)	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE
137		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
138		ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP
139	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE
140		TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR
141	PRO(.97)	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
142		GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
143		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
144	VAL(.95)	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL
145		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	LYS	THR	THR	THR	THR	THR	THR	THR	THR
146	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL
		---	---	---	---	---	---	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
		---	---	---	---	---	---	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
147		ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
148	TRP	TRP	TRP	TRP		TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP		TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP
149		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
150		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	ALA	GLY	ALA	ALA	---	ALA	ALA	ALA
		---	---	---	---	---	---	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
		---	---	---	---	---	---	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
151		ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ASP		ASP	ASP	ASP	ASP	ASP	---	ASP	ASP	ASP
152		SER	SER	SER	GLY	SER	SER	GLY	SER	SER	SER	SER	GLY	SER		GLY	SER	SER	SER	GLY	---	SER	SER	SER
		---	---	---	---	---	---	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
153		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
154		PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	---	PRO	PRO	PRO
155		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	ILE	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	---	VAL	VAL	VAL
156		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		ASN	LYS	LYS	LYS	LYS	---	LYS	LYS	LYS
157		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA		THR	ALA	ALA	ALA	ALA	---	ALA	ALA	ALA
158		GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY		GLY	GLY	GLY	GLY	GLY	---	GLY	GLY	GLY
159		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	---	VAL	VAL	VAL
160		GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	---	GLU	GLU	GLU
161		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	---	THR	THR	THR
162		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	---	THR	THR	THR
163		THR	THR	THR	LYS	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR	THR	THR	THR	LYS	---	THR	THR	ALA
164		PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	---	PRO	PRO	PRO
165		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
166	LYS(.98)	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	---	LYS	LYS	LYS
167		GLN	GLN	GLN	GLN	GLN		GLN	GLN	GLN	GLN	GLN	GLN	GLN		GLN	GLN	GLN	GLN	GLN	---	GLN	GLN	GLN
168		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
169		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
170	ASN(.96)	ASN	ASN	ASN	ASN	ASN		ASN	ASN	ASN	ASN	ASN	ASN	ASX		ASN	ASN	ASN	ASN	ASN	---	ASN	ASN	ASN
		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
171		ASN	ASN	ASN	ASN	ASN		ASN	ASN	ASN	ASN	ASN	ASN	ASX		ASN	ASN	ASN	ASN	ASN	---	ASN	ASN	ASN
172		LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	---	LYS	LYS	ARG
173		TYR	TYR	TYR	TYR	TYR		TYR	TYR	TYR	TYR	TYR	TYR	TYR		TYR	TYR	TYR	TYR	TYR	---	TYR	TYR	TYR
174		ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	---	ALA	ALA	ALA
175	ALA(.98)	ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	---	ALA	ALA	ALA
176	SER(.97)	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
177	SER(.98)	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
178		TYR	TYR	TYR	TYR	TYR		TYR	TYR	TYR	TYR	TYR	TYR	TYR		TYR	TYR	TYR	TYR	TYR	---	TYR	TYR	TYR
179	LEU	LEU	LEU	LEU		LEU		LEU	LEU	LEU	LEU	LEU	LEU	LEU		LEU	LEU	LEU	LEU	LEU	---	LEU	LEU	LEU
180		SER	SER	SER		SER		SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
181	LEU(.98)	LEU	LEU	LEU		LEU		LEU	LEU	LEU	LEU	LEU	LEU	LEU		LEU	LEU	LEU	LEU	LEU	---	LEU	LEU	LEU
182		THR	THR	THR		THR		THR	THR	THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	---	THR	THR	THR
183		PRO	PRO	PRO		PRO		PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	---	PRO	PRO	PRO
184		GLU	GLU	GLU		GLU		GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLX	GLU	---	GLU	GLU	GLU
185		GLN	GLN	GLN		GLN		GLN	GLN	GLN	GLN	GLN	GLN	GLN		GLN	GLN	GLN	GLX	GLN	---	GLN	GLN	GLN
186	TRP(.98)	TRP	TRP	TRP		TRP		TRP	TRP	TRP	TRP	TRP	TRP	TRP		TRP	TRP	TRP	TRP	TRP	---	TRP	TRP	TRP
187		LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	---	LYS	LYS	LYS
		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---
188		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	---	SER	SER	SER
189		HIS	HIS	HIS	HIS	HIS		HIS	HIS	HIS	HIS	HIS	HIS	HIS		HIS	HIS	HIS	HIS	HIS	---	HIS	HIS	HIS
190		LYS	ARG	ARG	ARG	ARG		ARG	ARG	LYS	ARG	ARG	ARG	ARG		ARG	ARG	ARG	ARG	ARG	---	LYS	LYS	ARG
191		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	---	SER	SER	SER
192		TYR	TYR	TYR	TYR	TYR		TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	---	TYR	TYR	TYR
193		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	---	SER	SER	SER
194	CYS	CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	---	CYS	CYS	CYS
195		GLX	GLN	GLU	GLN	GLN		GLN	GLN	GLN	GLU	GLN	GLN	LEU	GLN	GLN	GLN	GLN	GLX	GLN	---	GLN	GLU	GLN
196	VAL	VAL	VAL	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	---	VAL	VAL	VAL
197		THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	---	THR	THR	THR
198	HIS	HIS	HIS	HIS	HIS	HIS		HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	---	HIS	HIS	HIS
199		GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLX	GLU	---	GLU	GLU	GLU
200		GLY	GLY	GLY	GLY	GLY		GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	---	GLY	GLY	GLY
201		---	---	---	---	---		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
202		---	---	---	---	---		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
203		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	---	SER	SER	SER
204		THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	---	THR	THR	THR
205		VAL	VAL	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	---	VAL	VAL	VAL
206		GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLX	GLU	---	GLU	GLU	GLU
206A		---	---	---	---	---		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
207	LYS(.97)	LYS	LYS	LYS		LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	---	LYS	LYS	LYS
208		THR	THR	THR		THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	---	THR	THR	THR
209		VAL	VAL	VAL		VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	---	VAL	VAL	VAL
		---	---	---		---		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
		---	---	---		---		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
210		ALA	ALA	ALA		ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	---	ALA	ALA	ALA
211		PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	---	PRO	PRO	PRO
212		THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR		ALA	THR	THR	THR	THR	---	THR	THR	THR
213		GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	---	GLU	GLU	GLU
214	CYS	CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS	CYS	---	CYS	CYS	CYS
215		SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER		ALA	SER	SER	SER	SER	---	SER	SER	SER
216		---	---	---	---	---		---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---

LAMBDA LIGHT CONSTANT CHAINS (cont'd)

	24 KE-OZ- 'CL	25 KE-OZ+ 'CL	26 MCG 'CL	27 14.1 'CL	28 16.1 'CL	29 MOUSE L2'CL	30 MOUSE CL2S 'CL	31 1C4 D8.8 'CL	32 MOUSE L4'CL	33 MOUSE L1'CL	34 MOUSE L3'CL	35 MOPC 104E	36 PLAI -13 'CL #	37 IG 303 LAMBDA 'CL #	38 S43 'CL	39 MOPC 315	40 KL.SJA 83.12 'CL	41 MOPC 315A 'CL	42 6-2	43 MOPC 315-26 'CL'
107A	---	---	---	GLY	GLY	GLY		GLY	GLY	GLY	GLY	---	---	---	---	---	GLY	---		---
108	GLN	GLN	---	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN		GLN
109	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO
109A	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		---
110	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS
111	ALA	ALA	ALA	ALA	ALA	SER	SER	SER	ALA	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER
112	ALA	ALA	ASN	THR	THR	THR	THR	SER	THR	SER	THR	SER	SER	SER	SER	THR	SER	SER		THR
113	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO
114	SER	SER	THR	SER	SER	THR	THR	THR	SER	SER	THR	SER	SER	SER	SER	THR	SER	SER		THR
115	VAL	VAL	VAL	VAL	VAL	LEU	LEU	LEU	VAL	VAL	LEU	VAL	VAL	VAL	VAL	LEU	VAL	VAL		LEU
116	THR	THR	THR	THR	THR	THR	THR	THR	ASN	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
117	LEU	LEU	LEU	LEU	LEU	VAL	VAL	VAL	LEU	LEU	MET	LEU	LEU	LEU	LEU	VAL	LEU	LEU	MET	VAL
118	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE
119	PRO	PRO	PRO	PRO	LEU	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
120	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
121	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER
122	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	PRO	SER	SER	SER	SER	SER	SER	SER	PRO	SER
123	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU
124	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU
	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
124A	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
125	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
126	GLN	GLN	GLN	GLN	GLN	LYS	LYS	LYS	LYS	GLU	GLN	THR	GLU	THR	GLU	LYS	GLU	GLU	GLN	LYS
127	ALA	ALA	ALA	ALA	ALA	GLU	GLU	GLU	THR	THR	GLU	GLU	THR	GLU	THR	GLU	THR	THR	GLU	GLU
128	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	LYS	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN
129	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
130	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
131	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR
132	LEU	LEU	LEU	LEU	LEU	LEO	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
133	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL
134	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
135	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	MET	THR	LEU	THR	THR	THR	THR	LEU	THR	THR	LEU	LEU
136	ILE	ILE	ILE	MET	MET	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE	ILE
137	SER	SER	SER	ASN	ASN	SER	SER	SER	THR	THR	SER	THR	THR	THR	THR	SER	THR	THR		SER
138	ASP	ASP	ASP	ASP	ASP	ASN	ASN	ASN	GLU	ASP	ASN	ASP	ASP	ASP	ASP	ASN	ASP	ASP		ASN
139	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE		PHE
140	TYR	TYR	TYR	TYR	TYR	SER	SER	SER	TYR	TYR	SER	TYR	TYR	TYR	TYR	SER	TYR	TYR		SER
141	PRO	PRO	PRO	PHO	LEU	PRO	PRO	PRO	ALA	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO
142	GLY	GLY	GLY	GLY	GLY	SER	SER	SER	ALA	GLY	SER	GLY	GLY	GLY	GLY	GLY	GLY	GLY		SER
143	ALA	ALA	ALA	ILE	ILE	GLY	GLY	VAL	ALA	VAL	GLY	VAL	VAL	VAL	VAL	SER	VAL	VAL		GLY
144	VAL	VAL	VAL	LEU	LEU	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL
145	THR	THR	THR	THR	THR	THR	THR		ARG	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
146	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
147	ALA	ALA	ALA	THR	THR	ALA	ALA		ALA	ASP	ALA	ASP	ASP	ASP	ASP	ALA	ASP	ASP		ALA
148	TRP	TRP	TRP	TRP	TRP	TRP	TRP		TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP		TRP
149	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS
150	ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	VAL	ALA	VAL	VAL	VAL	VAL	ALA	VAL	VAL		ALA
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
151	ASP	ASP	ASP	ASP	ASP	ASN	ASN		ASP	ASP	ASN	ASP	ASP	ASP	ASP	ASN	ASP	ASP		ASN
152	SER	SER	GLY	GLY	GLY	GLY	GLY		GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	VAL	GLY		GLY
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
153	SER	SER	SER	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
154	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO
155	VAL	VAL	VAL	ILE	ILE	ILE	ILE		PHE	VAL	ILE	VAL	VAL	VAL	VAL	ILE	VAL	VAL		ILE
156	LYS	LYS	LYS	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
157	ALA	ALA	ALA	GLN	GLN	GLN	GLN		GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN		GLN
158	GLY	GLY	GLY	GLY	GLY	GLY	GLY		GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY		GLY
159	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL	MET	VAL	MET	MET	MET	MET	VAL	MET	MET		VAL
160	GLU	GLU	GLU	GLU	GLU	ASP	ASP		GLU	GLU	ASP	GLU	GLU	GLU	GLU	ASP	GLU	GLU		ASP
161	THR	THR	THR	MET	MET	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
162	THR	THR	THR	THR	THR	SER	SER		THR	THR	SER	THR	THR	THR	THR	SER	THR	THR		SER
163	THR	THR	LYS	THR	THR	ASN	ASN		GLN	GLN	ASN	GLU	GLN	GLU	GLN	ASN	GLN	GLN		ASN
164	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO		PRO
165	SER	SER	SER	SER	SER	THR	THR		PRO	SER	THR	SER	SER	SER	SER	THR	SER	SER		THR
166	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS
167	GLN	GLN	GLN	GLN	GLN	GLU	GLU		GLN	GLN	GLU	GLN	GLN	GLN	GLN	GLU	GLN	GLN		GLU
168	SER	SER	SER	SER	SER	GLY	GLY		ARG	SER	ASP	SER	SER	SER	SER	GLY	SER	SER		GLY
169	---	---	---	---	---	---	---		ARG	---	---	---	---	---	---	---	---	---		---
170	ASN	ASN	ASN	ASN	ASN	---	---		ASN	ASN	---	ASN	ASN	ASN	ASN	ASN	ASN	ASN		---
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
171	ASN	ASN	ASN	ASN	SER	ASN	ASN		---	ASN	ASN	ASN	ASN	ASN	ASN	---	ASN	ASN		ASN
172	LYS	LYS	LYS	LYS	LYS	LYS	LYS		---	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS
173	TYR	TYR	TYR	TYR	TYR	PHE	PHE		---	TYR	TYR	TYR	TYR	TYR	TYR	PHE	TYR	TYR		PHE
174	ALA	ALA	ALA	ALA	MET	MET	MET		MET	MET	MET	MET	MET	MET	MET	MET	MET	MET	MET	MET
175	ALA	ALA	ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
176	SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
177	SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER
178	TYR	TYR	TYR	TYR	TYR	PHE	PHE		TYR	TYR	PHE	TYR	TYR	TYR	TYR	PHE	TYR	TYR	PHE	PHE
179	LEU	LEU	LEU	LEU	LEU	LEU	LEU		LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
180	SER	SER	SER	SER	SER	HIS	HIS		LEU	THR	HIS	THR	THR	THR	THR	HIS	THR	THR	HIS	HIS
181	LEU	LEU	LEU	LEU	LEU	LEU	LEU		PHE	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
182	THR	THR	THR	THR	THR	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR	THR
183	PRO	PRO	PRO	PRO	PRO	SER	SER		ALA	ALA	SER	ARG	ALA	ALA	ALA	SER	ALA	ALA	SER	SER
184	GLU	GLU	GLU	GLU	GLU	ASP	ASP		GLU	ARG	ASP	ALA	ARG	ARG	ARG	ASP	ARG	ARG	ASP	ASP
185	GLN	GLN	GLN	GLN	GLN	GLN	GLN		ALA	ALA	GLN	TYR	ALA	ALA	ALA	GLN	ALA	ALA	GLN	GLN
186	TRD	TRP	TRP	TRP	TRP	TRP	TRP		TRP	TRP	TRP	GLU	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP
187	LYS	LYS	LYS	ARG	ARG	ARG	ARG		GLU	GLU	ARG	ARG	GLU	GLU	GLU	ARG	GLU	GLU		ARG
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
188	SER	SER	SER	SER	SER	SER	SER		SER	ARG	SER	SER	ARG	ARG	ARG	SER	ARG	ARG		SER
189	HIS	HIS	HIS	ARG	ARG	HIS	HIS		HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS
190	ARG	LYS	ARG	ARG	ARG	ASN	ASN		ASP	SER	ASN	SER	SER	SER	SER	ASN	SER	SER	ASN	ASN
191	SER	SER	SER	SER	SER	SER	SER		SER	SER	SER	TYR	SER	SER	SER	SER	SER	SER		SER
192	TYR	TYR	TYR	TYR	TYR	PHE	PHE		TYR	TYR	PHE	SER	TYR	TYR	TYR	PHE	TYR	TYR	PHE	PHE
193	SER	SER	SER	SER	SER	THR	THR		SER	SER	THR	SER	SER	SER	SER	THR	SER	SER		THR
194	CYS	CYS	CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
195	GLN	GLN	GLN	GLN	GLN	GLN	GLN		HIS	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN		GLN
196	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL
197	THR	THR	THR	MET	MET	THR	THR		THR	THR	THR	THR	THR	THR	THR	THR	THR	THR		THR
198	HIS	HIS	HIS	HIS	HIS	HIS	HIS		HIS	HIS	HIS		HIS	HIS	HIS	HIS	HIS	HIS		HIS
199	GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU	GLU	GLU		GLU	GLU	GLU	GLU	GLU	GLU		GLU
200	GLY	GLY	GLY	GLY	GLY	GLY	GLY		GLY	GLY	GLY		GLY	GLY	GLY	GLY	GLY	GLY		GLY
201	---	---	---	---	---	---	---		---	---	---		---	---	---	---	---	---		---
202	---	---	---	---	---	---	---		---	---	---		---	---	---	---	---	---		---
203	SER	SER	SER	SER	SER	ASP	ASP		GLU	HIS	ASP		HIS	HIS	HIS	ASP	HIS	HIS		ASP
204	THR	THR	THR	THR	THR	THR	THR		HIS	THR	THR		THR	THR	THR	THR	THR	THR		THR
205	VAL	VAL	VAL	VAL	ALA	VAL	VAL		VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL	VAL		VAL
206	GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLY	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU
206A	---	---	---	---	---	---	---		---	---	---	---	---	---	---	---	---	---		---
207	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS		LYS
208	THR	THR	THR	THR	THR	SER	SER		SER	SER	SER	SER	SER	SER	SER	SER	SER	SER		SER
209	VAL	VAL	VAL	VAL	VAL	LEU	LEU		LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU		LEU
	---	---	---	---	---	---	---		---	---	---	---	---		---	---	---	---		---
	---	---	---	---	---	---	---		---	---	---	---	---		---	---	---	---		---
210	ALA	ALA	ALA	ALA	ALA	SER	SER		SER	SER	SER	SER	SER		SER	SER	SER	SER		SER
211	PRO	PRO	PRO	PRO	PRO	PRO	PRO		ARG	ARG	PRO	ARG	ARG		ARG	PRO	ARG	ARG		PRO
212	THR	THR	THR	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA	ALA		ALA	ALA	ALA	ALA		ALA
213	GLU	GLU	GLU	GLU	GLU	GLU	GLU		GLU	ASP	GLU	ASP	ASP		ASP	GLU	ASP	ASP		GLU
214	CYS	CYS	CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS	CYS		CYS	CYS	CYS	CYS		CYS
215	SER	SER	SER	SER	SER	LEU	LEU		SER	SER	LEU	SER	SER		SER	LEU	SER	SER		LEU
216	---	---	---	---	---	---	---		---	---	---	---	---		---	---	---	---		---

ELSEVIER

Journal of Immunological Methods 247 (2001) 119–130

JIM Journal of Immunological Methods

www.elsevier.nl/locate/jim

Baculovirus expression cassette vectors for rapid production of complete human IgG from phage display selected antibody fragments

Mifang Liang[a,*], Stefan Dübel[b], Dexin Li[a], Iris Queitsch[b], Wei Li[a], Ekkehard K.F. Bautz[b,c]

[a]*Institute of Virology, Chinese Academy of Preventive Medicines, 100 Ying Xin Jie, Xuan Wu Qu, Beijing 100052, China*
[b]*Institute of Molecular Genetics, University of Heidelberg, Im Neuenheimer Feld 230, D-69120, Heidelberg, Germany*
[c]*Hantavirus-Forschungsstelle der Heidelberger Akademie der Wissenschaften, Im Neuenheimer Feld 230, D-69120, Heidelberg, Germany*

Received 7 July 2000; received in revised form 17 October 2000; accepted 19 October 2000

Abstract

For the expression of human intact IgG antibodies, we have constructed a set of baculovirus expression vectors designed to facilitate rapid insertion of heavy and light chain genes of Fab or scFv antibodies derived from phage display antibody libraries. By linking them to human constant or Fc regions, expression of complete human immunoglobulin molecules was achieved in insect cells by infection with recombinant baculovirus. The IgG expression cassette vectors are based on the backbone vector which contains two back to back polyhedron and p10 promoters. The IgG expression cassette elements, including the authentic IgG lambda or kappa and heavy chain signal sequences, as well as light chain (lambda or kappa) and heavy chain constant region genes are combined in a single vector and are controlled by the p10 and polyhedron promoter respectively. Either of VL or Fab-L and VH or Fab-Fd genes from common phage display systems can be directly inserted into one of the cassette vectors through in-frame cloning sites. This design of a single cassette vector combining heavy and light chain expression elements allowed rapid production and secretion of correctly processed and assembled intact immunoglobulins from recombinant baculovirus infected insect cells. The recombinant antibodies showed the expected molecular size of the H2L2 heterodimer in non reducing SDS–PAGE. No apparent differences were found between the expression level of heavy and light chains, and antigen binding function was preserved. For various antibodies, yields between 6 and 18 mg/l IgG were obtained.

Keywords: Cassette vector; Complete IgG antibody; Baculovirus expression; Phage display; Antibody engineering

1. Introduction

Within the past decade, antibody phage display technology has been established as a proven technology to select scFv or Fab antibody fragments specific for various antigens. The phage display technology

*Corresponding author. Tel.: +86-10-635-81325; fax: +86-10-635-32053.
E-mail address: liangmf@public3.bta.net.cn (M. Liang).

0022-1759/01/$ – see front matter
PII: S0022-1759(00)00322-7

EXC

has particularily boosted the generation of human antibody fragments. Human antibodies do not induce an immune response against the antibody when applied to the human body. Therefore, human IgG molecules are of particular value for therapy and in vivo diagnostics, but cannot be easily obtained by non-recombinant methods. *E. coli* cells, despite being optimal for the recombinant antibody phage selection procedure, are on the other hand not suitable to produce complete functional IgG molecules, since they cannot provide appropriate folding, disulfide bond formation and several post-translational modifications. The phage selection process yields the antigen binding regions only, either in form of scFv or Fab antibody fragments, which have to be converted to the complete diheterotrameric IgG molecule. Since the phage display selections usually result in a number of candidate antibody fragments, and functional tests frequently require complete IgG, a eukaryotic expression system is required for fast and convenient production.

The baculovirus expression system has been applied widely for the expression and study of genes and their proteins from various sources. Furthermore, it has already been established as a reliable system for the production of complete chimeric, humanized or human IgG that are similar to native molecules both structurally and functionally (Nesbit et al., 1992; Poul et al., 1995a,b; Liang et al., 1997a; Porter et al., 1997; Tan and Lam, 1999). In contrast to the quite time consuming establishment of a stable mammalian IgG expression cells lines, the baculovirus expression system offers obvious advantages by saving time and allowing rapid scale up of production.

Cassette transfer vectors for the expression and secretion of intact human antibodies have been reported previously (Poul et al., 1995a,b). However, these systems require a combination of two vectors which separately served for VH and VL cloning, requiring careful and time consuming adjustment of titers of the two respective recombinant baculoviruses. We present here a set of single, universal baculovirus human or humanized antibody expression vectors with authentic IgG light and heavy chains signal sequences, which were specifically designed for direct cloning of the heavy and light genes of Fab or scFv antibodies selected from phage display libraries. By linking them to a human constant region, a complete IgG is expressed and secreted by insect cells after infection with a single recombinant baculovirus clone. The functionality of these vectors has been verified for various VH/VL or Fab genes obtained from phage display libraries.

2. Materials and methods

2.1. Materials

Cloning vectors PCR™II and pUC18, pGEM5zf were purchased from Invitrogen (Groningen, Netherlands) or Pomega (Mannheim, Germany). Backbone vector pACUW51, Sf9 cells and H5 cells were purchased from Pharmingen (Heidelberg, Germany). *E. coli* DH5r used for cloning was purchased from (GibcoBRL, Karlsruhe, Germany). FITC or HRP conjugated anti-human Fab, human Kappa, lambda and Fc antibodies were purchased from Sigma (München, Germany). The hantaan virus vero-E6 antigen slides were provided by Progen (Heidelberg, Germany). The human VH, VL or Fab genes were isolated in our lab or provided by the Institute of Virology, Chinese Academy of Preventive Medicine, Beijing, China. Human standard IgG for standardization of the ELISA was purchased from Sigma (München, Germany).

2.2. PCR amplification of the genetic elements for the IgG expression vector cassette

All standard cloning procedures were carried out as described by Sambrook et al. (1989). Oligonucleotide primers used for the PCR amplification of heavy and light chain genes of IgG expression cassette are listed in Table 1. Total cellular RNA was prepared from pelleted human lymphocytes using Trizol reagent (Gibco/BRL, USA). cDNA was synthesized using oligdT primers and reverse transcriptase (Gibco/BRL, USA). The full length IgG1 heavy chain gene was amplified with primers IgG NSVH3 and IgG CH3 (Table 1). Thirty-five PCR cycles were performed, with incubations for 1 min at 94°C, 1 min at 54°C and 3 min at 72°C each. The 69 base pairs oligonucleotide of kappa and lambda leader sequence DNAs containing mutated cloning sites SacI and

Table 1
PCR primers used fo rthe construction of the human IgG expression cassette vectors

Name	Oligonucleotide sequences	Cloning sites
NSVH3F	5′-CGC*GGATCC*ACCATGGAGTTTGGGCTGAGC-3′	BamH1
CH3R	5′-CGC*GGATCC*CTCCTGCGTGTAGTGGTTGTGC-3′	BamH1
VKNSF	5′-GGA*AGATCT*CACCATGGAAACCCCAGCGCA-3′	BglII
VKNSR	5′-GGA*AGATCTGATATCT*CGCGT*GAGCT*CAATTTCTTCGCTGCTT-3′	BglII, EcoRV, SacI
VLNSF	5′-CGAA*GATCT*AGCATGGCCTGGTCTCCTCTC-3′	Bgl II
VLNSR	5′-GGA*AGATCTGATATC*CGGCT*GGAGCTC*CACAGACTGGG-3′	BglII, EcoRV, SacI
VK-Hind-F	5′-ACA*CTTGAGCT*CAGGGGACC*AAGCTT*GAGATC-3′	SacI, HindIII
CK-EcoR-R	5′-CCG*GATATC*TAGAACTAACACTCTCCCCTGTTGA-3′	EcoRV
VK-Hind-F	5′-TGGGTT*GAGCTC*GGAGGGACC*AAGCTT*ACCGTC-3′	SacI, HindIII
CK-EcoR-R	5′-CCG*GATATC*TAGAACTATGAACATTCTGTAGG-3′	EcoRV
VH3-NS-R	5′-CCCAGA*CTCGAG*CAGTTGCACCTC-3′	XhoI
CH1-XN-F	5′-CCG*CTCGAG*CGTCTCCTCA*GCTAGC*ACCAAGGGCCCATC-3′	XhoI, NheI
CH2-XS-F	5′-CCG*CTCGAG*CGGTGACAAAA*CTAGT*ACATGCCCACCGTGCC-3′	XhoI,SpeI
pAc-Bam-F	5′-CCTATAAATA*CGGATC*CGGTTAT-3′	BamH1
HindMuF	5′-CGTAAACACGTTAAATA*GAGCTT*GGACA-3′	HindIII-Mutant
HindMuR	5′-TGTC*CAAGCTC*TATTTAACGTGTTTACG-3′	HindIII-Mutant
PAc-Sca-R	5′-TGACTGGTGAGTACTCAACCAAGT-3′	SacI

EcoRI were amplified with primers VKNSF and VKNSR or VLNSF and VLNSR, respectively, by 30 PCR cycles at 94°C min for 1 min, 54°C for 1 min and 72°C for 1 min. With Primers VK-Hind-F and CK-EcoR-R or VL-Hind-F and CL-EcoR-R, the constant region genes of kappa or lambda chains containing the VL cloning sites SacI and HindIII were obtained by a similar PCR amplification of the kappa or lambda genes of two human derived Fab antibodies (Liang et al., 1997b). All of the amplified PCR fragments were purified by using Qiagen Gel-extraction kit (Qiagen, Hilden, Germany) and stored at −20° for subsequent cloning.

2.3. Construction of recombinant baculovirus IgG expression cassette vectors

Four recombinant baculovirus IgG expression vectors (pAc-K-CH3, pAc-L-CH3,pAc-K-Fc and pAc-L-Fc) with different combinations of the essential genetic elements for heavy and light chain expression have been constructed as follows The PCR product of the full length human heavy chain DNA including the authentic IgG1 heavy chain leader DNA were cloned into the BamHI site of plasmid pUC18. To remove the gene fragments encoding for the antigen binding parts, the fragments containing the modified heavy chain gene with leader sequence and complete constant region or Fc portion, were PCR amplified with the primer sets VH3-NS-R and CH1-XN-F or VH3-NS-R and CH2-XS-F (Table 1) from two opposite ends using the PUC18-heavy chain vector as a template. The resulting PCR products omitted the VH or VH-CH1 region of the original heavy chain fragment, they were self ligated and transformed into *E. coli* DH5r competent cells. Above steps resulted in two transfer vectors containing heavy chain expression cassettes: 1.PUC-H-NheI, which contains the IgG1 heavy chain leader sequnece, the VH in-frame cloning sites XhoI and NheI and the complete constant region; and 2. PUC-H-SpeI, which contains the same leader sequence, the Fd in-frame cloning sites XhoI and SpeI and the Fc region.

To obtain the vector pAc-K-CH3 or pAc-L-CH3, the original HindIII site in the backbone vector pAcUW51 was mutated by assembly PCR with primers pAc-Bam-F, HindMuF, HindMuR, pAc-Sca-R (Table 1). In this step, the HindIII site was replaced by the sequence GAGTTC. The purified PCR products encoding kappa or lambda signal sequences followed by VL-CL fragment cloning site SacI and EcoR1 were ligated into the BglII site of the backbone vector pAcUW51, resulting in two transfer vectors: pAc-K-Leader and pAc-L-leader. The constant region fragment genes of kappa or lambda chains were then cloned into the above transfer vectors pAc-K-leader or pAc-L-leader

through SacI and EcoRI, resulting in two new transfer vectors: pAc-K-CK and pAc-L-CL. The heavy chain expression cassette was cleaved from the vector PUC-H-NheI with BamHI and cloned into pAc-K-CK or pAc-L-CL vectors, to finally yield the two expression vectors pAc-K-CH3 and pAc-L-CH3. In a similar way, the DNA fragment of the heavy chain expression cassette was obtained from the respective vector PUC-H-SpeI and cloned into the BamHI site of pAc-K-leader or pAc-L-leader, finally to yield the vectors pAc-K-Fc and pAc-L-Fc. All mutations, insertions or deletions of the resulting vectors were controlled by DNA sequencing by using the ABI auto-sequence kit and the ABI 310 automated capillary sequencer (Perkin Elmer, Langen, Germany).

2.4. Insertion of V region or Fab genes

Various human or mouse recombinant IgG Fab fragment genes derived from phage display antibody libraries or mouse hybridoma cells were used to test the efficiency of the universal vector system. According to the sequence information of each individual antibody genes, the PCR amplification of the variable region of human or mouse IgG Fab antibody genes were performed with forward primers reported by Kang et al. (1991) and the following reverse primers: VH-NheIR:5′-TGG GCC CTT GGT GCT AGC TGA GGA GAC GGT GACC-3′; VL-HindIIIR: 5′-GAC GGT AAG CTT GGT CCC TCC-3′; VK-HindIIIR: 5′-GGATCTCAAGCTTGGT-CCCCT-3′; Mu-HindIIIR1: 5′-CAG CTC CAA GCT TGG TCC CAC CAC CGAA-3 and Mu-HindIIIR2 5′-TT CAG CTC AAG CTT GGT CCC GAA CG-3′. The PCR products were digested with XhoI and NheI (from heavy chain DNA) and SacI and HindIII (from light chain DNA) and cloned into the vector pAc-K-CH3 or pAc-L CH3 in accordance to the original kappa or lambda type. The Fab genes inserted into vectors pAc-K-Fc or pAc-L Fc were directly cleaved from the pComb3 phagemid based Fab clones (Barbas et al., 1991) with XhoI and SpeI or SacI and XbaI. The respective light chain genes were subcloned into transfer vector pEG5F (Promega, Mannheim, Germany), and subsequently cloned into the SacI and EcoRV sites of above two vectors.

2.5. Preparation of recombinant baculovirus for the expression of intact human or mouse/human chimeric IgG

Recombinant baculoviruses were prepared by homologous recombination using the baculo–Gold transfection kit (Phamingen, Heidelberg, Germany) according to the instructions given by the supplier. Recombinant baculovirus was harvested 4–5 days after transfection from supernatants of SF9 cells culture medium, and subsequent plaque purification was performed to obtain pure and high titer recombinant virus. Intracellular heavy and light chain expression in the insect cells was tested by immunoflorescence using FITC conjugated anti-human Fc and anti-human Fab antibodies. Secreted recombinant human IgG antibodies were detected in the supernatants of infected SF9 cells by a conventional capature ELISA, using goat anti-human IgG Fab as a capture reagent and HRP conjugated anti-human Fc for detection.

2.6. Expression and purification of baculovirus expressed IgG antibody

SF9 cells or H5 cells were infected with the recombinant viruses expressing various IgG antibodies at an m.o.i of 10, and grown in serum-free medium (GibcoBRL, Karlsruhe, Germany), incubated in T75 flasks, at 27° until approximately 50–60% of dead cells were observed (approx. 4–5 days postinfection). The supernatants of recombinant baculovirus infected insect cells were harvested and clarified by centrifugation, filtered through 0.45 μm filters and applied to a Protein G-Sepharose CL-4B (Pharmacia, Braunschweig. Germany). The IgG fraction was eluted with 1.0 M glycine-HCL, pH 2.7, and neutralized with 1 M Tris, then applied to a desalting column eluted with 0.02 M sodium phosphate, pH 7.0. IgG concentrations were estimated according to Harlow and Lane (1988), and ajusted to a concentration of 200–500 ug/ml.

2.7. SDS–PAGE and western-blot

All expressed and purified IgG samples were analyzed under reducing and non-reducing condition on 10% polyacrylamide SDS–gels. SDS–PAGE was

carried out according to Lämmli (1970). Total protein staining was achieved with Coomassie Brilliant Blue R250 (Serva, Heidelberg, Germany). Immunoblots were performed essentially according to Towbin et al. (1980) and blocked with PBS–5% milk for 2–3 h before incubation with a mixture of HRP conjugated goat anti-human IgG and HRP conjugated goat anti human kappa+lambda chain antibodies. TMB Stabilized Substrate for HRP (Promega, Madison, USA) was used for visualizing bound enzymatic activity.

2.8. Immunoflorescence

For the detection of IgG antibodies expressed in SF9 cells, the cells cultured in T25 flasks were infected with recombinant viruses for 3–4 days. The cells were washed and suspended in PBS prior to fixation on 10 well slides with acetone at room temperature for 10 min. FITC conjugated anti-human Fab and Fc antibodies were subsequently incubated with the slides at 37°C for 30 min and washed 3 times with PBS prior to embedding for microscopic examination. To determine the specificity of the expressed IgG antibody, the immunofluorescence test for hantavirus infection of Progen (Heidelberg, Germany) was used. The supernatants of recombinant baculovirus infected SF9 cells containing IgG antibodies to hantavirus nucleocapsid protein (Liang et al., 1997b) were used as control. Bound antibodies were detected with FITC conjugated anti-human Fab and human Fc as described above.

2.9. Quantitative IgG ELISA

The Sf9 cells or H5 cells in T75 flasks were infected with 10^7 baculoviruses and incubated at 27° for 4–5 days until to 50–60% of the cells have died. The infected supernatants were harvested and tested by sandwhich ELISA. Commercialy available purified human IgG was used for a calibration curve by preparing a two fold dilution series in PBS. ELISAs were performed by coating anti-human Fab antibody onto a 96-well ELISA plate at 4°C overnight. Then, the wells were incubated with culture supernatants or the purified IgG dilutions. All washings were done with PBS. Bound immunoglobulins were detected by incubation with HRP conjugated anti human IgG Fc antibody and visualised using soluble TMB substrate (Promega, Madison, USA)). The O.D values at 450 nm of each sample was compared to the IgG calibration curve to calculate the protein concentration.

3. Results

3.1. Design and construction of baculovirus human IgG expression cassette vectors

A set of universal baculovirus expression cassette vectors (pAc-K-CH3, pAc-L-CH3, pAc-K-Fc and pAc-L-Fc) for the production of human IgG has been designed and constructed as illustrated in Fig. 1. The cassette vectors were specifically designed for direct insertion of the antibody Fv or Fab genes selected from phage display libraries and linking them to human IgG constant region, resulting in intact IgG expression vectors with heavy and light chain expression elements in one construct. The heavy chain elements are under control of the polyhedrin promoter, followed by the the 66 bp authentic IgG signal sequence DNA from IgG1 subgroup VHIII family and by the mutated in-frame cloning sites XhoI and NheI for cloning scFv VH genes (Fig. 1A) or XhoI and SpeI for cloning Fab Fd genes (Fig. 1B). The whole constant region gene of human IgG1 (Fig. 1A) or the Fc region gene of human IgG1(Fig. 1B) are located further downstream of the signal sequence DNA and these cloning sites. Having the opposite orientation compared to the heavy chain operon, in the same vector, the light chain elements are under the control of P10 promoter, followed by the 69 bp authentic signal sequence DNA of human kappa or lambda chain (Fig. 1A,B). This is followed by the mutated in-frame cloning site SacI and HindIII for cloning of V lambda or V Kappa region genes; followed by the constant region genes of Lambda or Kappa chains (Fig. 1A). The original HindII site in the vector pACUW51 was mutated in both vectors pAc-K-CH3 or pAc-L-CH3. In the vectors pAc-K-Fc and pAc-L-Fc, the cloning sites SacI and EcoRV were introduced adjacent to the signal sequences to allow cloning of Kappa or Lambda chain Fab genes, (Fig. 1B).

As the cassette vectors were designed with the

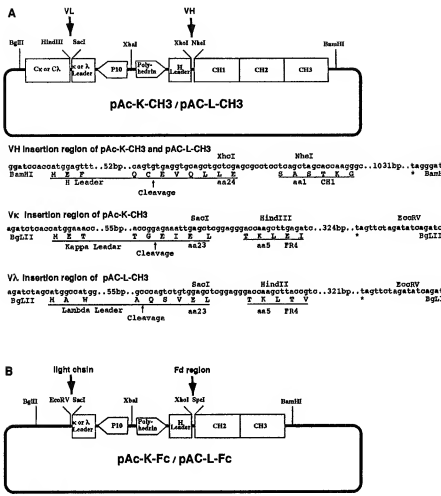


Fig. 1. Cassette baculovirus vectors for human IgG expression. A, vectors for the cloning of V region DNA obtained from scFv phage display systems; B, vectors for the cloning of Fab region DNA obtained from Fab phage display systems. The vector backbone in all cases (outside Bgl II/BamHI) is pAcUW51. Abbreviations: aa, amino acid number; C(x), constant human immunoglobulin regions; H, K or L leader, sequence coding for the respective human immunoglobuline signal peptides. The genetic elements are not drawn to scale.

purpose of cloning different forms of antibody genes from phage display libraries, the insertion sites for cloning of V genes in the vectors were chosen on the basis of their low cutting frequencies in human antibody V region genes (Persic et al., 1997) and the concern of the amino acid sequence conservation at the sites. The light and heavy chain cloning sites of the pComb3 phage display system (Barbas et al., 1991) were introduced into the vectors. Therefore, Fab genes from this phagemid could be directly cloned into the vectors pAc-K-Fc or pAc-L-Fc. Other forms of antibody fragments derived from different sources including human or rodent phage display libraries (McCafferty et al., 1990; Pope et al., 1996; Breitling and Dübel, 1997) could be inserted after a PCR to introduce the respective sites. As Fig. 1 indicated, the heavy chain 5′ cloning site XhoI must start at 72 bp, at amino acid (aa) position 24 since ATG of the IgG1 gene amplified from mRNA, with CTC=Leu; the 3′ cloning site NheI must start at the first amino acid of the constant region, with GCT=Ala; SpeI must start at 318 bp (aa106 of constant region, with ACT=Thr). The kappa or lambda 5′ cloning site SacI must start at 69 bp (aa23) to retain the ATG of kappa or lambda genes from mRNA; the 3′ cloning site HindIII must start at 18 bp (aa6, the first bp of framework 4, with AAG=Lys). The mutations introduced in all four vectors have been confirmed by DNA sequencing.

3.2. Expression in insect cells of functional recombinant human IgG antibodies derived from antibody variable or Fab genes

To check the function of the new vectors, various V region or Fab antibody genes were cloned into the vectors, comprising the VH, Vκ and Vλ genes of antibodies against hantavirus, hepatitis A virus and rabies virus (unpublished data). VH fragments derived from display libraries were cloned into XhoI and NheI sites, and Vκ or Vλ were cloned into SacI and HindIII sites of the vectors pAc-K-CH3 or pAc-L-CH3, respectively, in accordance with the original chain type. Fab antibody genes obtained from phage libraries or mouse hybridomas were directly cloned into the XhoI and SpeI sites (Fd) and SacI and EcoRV sites of the vectors of pAc-K-fc or pAc-L-Fc.

After plaque purification of the respective recombinant baculoviruses, the IgG antibodies expressed in Sf9 cells and H5 cells were analysed after infection. The strong cytoplasmic flurescence of infected Sf9 insect cells detected with FITC conjugated anti-human Fc and Fab antibodies (Fig. 2) demonstrated

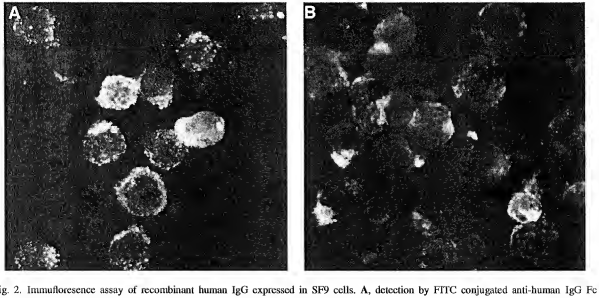


Fig. 2. Immunfloresence assay of recombinant human IgG expressed in SF9 cells. A, detection by FITC conjugated anti-human IgG Fc antibodies. B, detection by FITC conjugated anti-human IgG Fab antibodies.

that a human antibody Fab and Fc regions were well expressed.

To determine whether the recombinant human IgG immunoglobulins expressed either from pAc-K(L)-CH3 or pAc-K(L) vectors were correctly processed, assembled and secreted from the recombinant baculoviris infected insect cells, supernatants of cultures were collected. The recombinant immunoglobulins were recovered by binding to protein G-sepharose. Heavy and light chains of the expected sizes were observed under reducing condition of SDS–PAGE (Fig. 3a) and verified by incubating with HRP conjugated anti-human IgG Fc and anti human light chain antibodies on immunoblots (Fig. 3c). The purified recombinant human IgG molecules showed identical migrating size when compared to human blood derived IgG molecules under non-reducing conditions in SDS–PAGE electrophoresis (Fig. 3b). Similarily, bands with the expected molcular mass of IgG heavy and light chains were detected under reducing conditions. These results demonstrated that the recombinant human IgG antibodies produced in insect cells were correctly assembled and secreted into the recombinant baculovirus infected insect cell culture mediums as heterodimeric H2L2 immunoglobulins.

The expression levels provided by the universal IgG expression vectors were analyzed by using 5

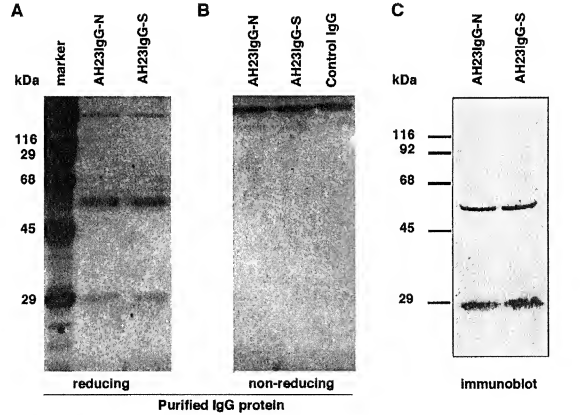


Fig. 3. SDS–PAGE and immunoblot analysis of recombinant human IgG produced by SF9 cells. **A, B**, coomassie blue stain, **C**, Immunoblot. Antibodies purified from the supernatants of recombinant baculovirus infected SF9 cells were analysed by SDS–PAGE using reducing (A) or nonreducing (B) conditions. Lane A-1, size Marker, lane A-2, IgG from vector pAc-L-CH3; lane A-3, IgG from vector pAc-L-Fc. Lane B-1, IgG from vector pAc-L-CH3, lane B-2, IgG from vector pAc-L-Fc. Lane B-3, Control (Human IgG franction obtained from Sigma, Deisenhofen, Germany). **C**, Immunoblot stainig with HRP conjugated anti-human IgG (Lane C-1, IgG from vector pAc-L-CH3, lane C-2, IgG from vector pAc-L-Fc).

human antibodies derived from phage antibody libraries which recognized hantavirus glycoprotein G1, hantavirus nucleocapsid protein (Liang et al., 1997b), hepatitis A VP1 proteins (Chao et al., 2000) and rabies virus (unpublished data). The ELISA to quantify the yield of secreted antibodies was standardized by comparing the titer of human IgG antibodies in the culture medium with human IgG immunoglobulin from blood donors. The expression levels of secreted human antibodies was around 6–18 mg/l (Fig. 4).

3.3. Functional analysis of the recombinant human antibodies expressed in insect cells

The function of the 5 baculovirus/insect cell expressed recombinant human antibodies was confirmed by specific binding to their related target antigens or by neutralizing target viruses. All IgGs showed the functional activity of their maternal antibody fragments. For example, we compared a human Fab fragent to hantavirus nuleocapsid protein obtained by phage display and the human IgG

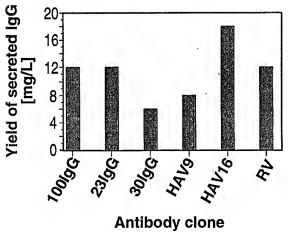


Fig. 4. Yields of recombinant human IgG antibodies secreted into the cell culture medium of High Five insect cells infected by recombinant baculoviruses. The production levels of various clones were determined by comparison to a calibration curve using human immunoglobulins of known concentration. Samples: 100IgG, R-IgG to Hantavirus G1 (vector pAc-L-CH3); 23IgG, R-IgG to Hantavirus G1 (vector pAc-L-Fc); 30IgG, R-IgG to Hantavirus NP (vector pAc-K-CH3); HAV9 and HAV16, R-IgGs to Hepatitis A (vector pAc-L-Fc); RV, R-IgG to Rabies virus (vector pAc-K-Fc).

antibody derived from it for their binding specificity on hantavirus infected Vero-E6 cells. The baculovirus expressed human IgG antibody showed strong binding (Fig. 5a) with the typical pattern obtained for intracellular virus stains, as demonstrated by the maternal Fab fragment (Fig. 5b).

4. Discussion

To achieve rapid expression of complete human immunoglobulins, we have designed and constructed a set of universal human antibody expression vectors with authentic IgG light and heavy chains signal sequences, mutant in-frame cloning sites and human IgG constant or Fc regions, which allow facile cloning of the heavy and light chain genes of Fab or scFv antibodies selected from phage display libraries or hybridomas and express them as intact IgG antibodies in recombinant baculovirus/insect cells system. The functions of the recombinant human (Fig. 4) or human-mouse chimeric antibodies (data not shown) with our cassette vector system were verified by the specific binding to their target antigens.

A quick procedure for expressing intact human IgG antibodies derived from Fab or scFv genes selected from phage display libraries has turned out to be essential for various methods of functional characterization, in particular in animal models. Prokaryotic expression systems offer the quickest solution for the expression of foreign genes; but they cannot produce and correctly fold the entire IgG diheterotetramer. This can only be achieved in eukaryotic systems. Mammalian cell expression systems have been widely used to produce functional IgG. However, the yield of transient expression systems is usually low, and to achieve stable expression requires time consuming selection procedures. Baculovirus/insect cells systems, in contrast, combine the advantages of time saving, correct protein folding and stable and large scale expression (Hasemann and Capra, 1990)

The intact IgG expression cassette vectors that we addressed in this paper are based on the backbone vector pAcUW51, which has been successfully used before to express an IgG to hantavirus glycoprotein G1 which retained the neutralizing activity of its

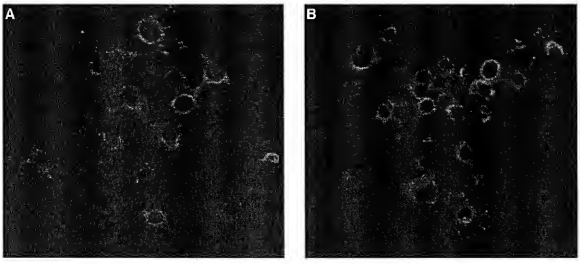


Fig. 5. Immunoflurescence staining of cells expressing hantavirus antigens by recombinant human antibodies produced in baculovirus/insect cells. Antigen slides with Hantavirus infected Vero-E6 cells were used. **A**, maternal Fab fragment produced in *E. coli*; **B**, corresponding human IgG produced in insect cells using the vector pAc-K-Fc.

maternal human hybridoma monoclonal antibody (Liang et al., 1997b). The universal cassette described in this study, however, by encoding both heavy and light chains under the control of different back-to-back promoters avoid the disadvantage of previous separate expression vector systems for H and L chain. These required significant efforts to adjust the expression levels of heavy and light chains by evaluating the optimal ratio of the different recombinant baculovirus titers, and it required two separate steps of baculovirus recombination. As these vectors will be used with the purpose of rapid and simple cloning of Fab or scFv antibody genes from phage display libraries, we paid particular attention to the design of the restriction cloning sites. Most of the introduced in-frame unique restriction sites (Xho I and NheI for the insertion of VH region and SacI and HindIII for the insertion of VL region) in the vectors pAc-K-CH3 or pAc-L-CH3 do not result in an amino acid exchange, except for the change to introduce the SacI site resulting in a mutation from a neutral valine to a negatively charged glutamic acid residue. Although some data has shown that FR1 residues can considerably influence antigen binding and antibody affinity (Xiang et al., 1991), previous investigations have demonstrated that the introduction of this particular mutation did not influence the antibody affinity (Bender et al., 1993). In the heavy chain cassette of the vectors pAc-K-Fc or pAc-L -Fc, the introduction of the SpeI site resulted in an amino acid change from histidine to serine at the beginning of FR2 region, thus decreasing the hydrophilicity of the amino acid at the site. This change, however, did not influence functionality of all three antiviral neutralizing antibodies tested in this study. For example, Fab genes of a neutralizing antibody to hepatitis A selected by phage display were directly cloned into the vector pAc-L-Fc, which contains the SacI mutation in the light chain and the SpeI mutation in the heavy chain gene. The resulting recombinant human IgG antibody maintained the specificity and showed a better neutralizing activity for hepatitis A virus and higher binding affinity, reaching picmolar dissociation constants (unpublished data). This increase of apparent affinity is in accordance with the increase of avidity expected from the bivalent binding of the IgG when compared to the monovalent Fab ragment. In gener-

al, we recommend that the light chain is cloned first, followed by the heavy chain, since the frequency of the restriction sites employed for heavy chain cloning are less abundant in light chain germline V regions.

For maximum authenticity of the human recombinant antibodies, we constructed separate vectors with either human kappa or lambda IgG signal sequences and correlated constant regions. In our experiments, there was no significant difference in yield between kappa or lambda expression cassette when the variable region of kappa or lambda was appropriately cloned. This, however, may not be a significant issue in the expression of full-length functional antibodies with different light chains. The yield of secreted antibody achieved in the presented baculovirus/insect cell system reached 6–18 mg/l and about 50–70% antibody protein could be recovered after protein G affinity purification. High Five insect cells have already been adapted for the large scale production in a stirred bioreactor (Savary et al., 1999), thus allowing continuous feed fermentation with the baculovirus system.

Acknowledgements

This work was supported by Chinese National Foundation of '863' R&D High Technology Programme and the Heidelberger Akademie der Wissenschaften, Forschungsstelle Hantaviren.

References

Barbas, III C.F., Kang, A.S., Lener, R.A., Nenkovic, S.J., 1991. Assembly of Combinatorial antibody libraries ob phage surfaces: The gene III site. Proc. Natl. Acad. Sci. USA 89, 4457.

Bender, E., Woof, J.M., Atkin, J.D., Barker, M.D., Bebbington, C.R., Burton, D., 1993. Recombinant human antibodies: linkage of an Fab fragment from a combinatorial library to an Fc fragment for expression in mammalian cell culture. Hum. Antibod. Hybridomas 4, 74.

Breitling, F., Dübel, S., 1997. Cloning and Expression of Single Chain Fragments (scFv) from Mouse and Rat Hybridomas. In: Reischl, U. (Ed.), Molecular Diagnosis of Infectious Diseases. Methods in Molecular Medicine, Vol. 13. Humana Press Inc, Totowa, NY.

Chao, J., Liang, M., Guo, K., Meng, Q., Shang, D., Li, D., Zhan, M. (2000) Human Fan antibodies derived from phage display library neutralize Hapatitis A virus. Chin. J. Exp. Clin. Virol., in press.

Harlow, E., Lane, D., 1998. Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory (CSH), NY.

Hasemann, C.A., Capra, J.D., 1990. High-level production of a functional immunoglobulin heterodimer in a baculovirus expression system. Proc. Natl. Acad. Sci. USA 87, 3942.

Kang, A.S., Burton, D., and Lerner, R.A. 1991. Combinatorial Immunoglobulin libraries in phage λ. Methods: A Companion to Meth. in Enzymol. 2, 111.

Lämmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nature 227, 680.

Liang, M., Guttieri, M., Ludviska, A., Schmaljohn, C.S., 1997a. Baculovirus expression of a human G2 specific neutralizing IgG monoclonal antibody to Puumala Virus. Virology 235, 1997.

Liang, M., Li, D., Hang, C., Wu, X., Zhu, G., Xue, Y., Li, C., Song, G., 1997b. Generation of human neutralizing monoclonal antibody Fab fragments to hanataam virus by using phage display technology. Chinese J. Virol. 13, 297.

McCafferty, J., Griffiths, A.D., Winter, G., Chiswell, D.J., 1990. Phage antibodies: filamentous phage displaying antibody variable domains. Nature 384, 552–554.

Nesbit, M., Fu, Z.F., McDonald-Smith, J., Steplewski, Z., Curtis, P.J., 1992. Production of a functional monoclonal antibody recognizing human colorectal carcinoma cells from a baculovirus expression system. J. Immunol. Meth. 151, 201.

Persic, L., Roberts, A., Wilton, J., Cattaneo, A., Bradbury, A., Hoogenboom, H.R., 1997. An intergrated vector system for the eukaryotic expression of antibodies or their fragments after selection from phage display libraries. Gene 187, 9.

Pope, A.R., Embleton, M.J., Memaugh, R., 1996. Construction and use of antibody gene reperoires. In: McCafferty, J., Hoogenboom, H.R., Chiswell, D.J. (Eds.), Antibody Engineering, A practical Approach. (PAS). IRL Press at Oxford University Press.

Porter, A.J.R., Bentley, K.J., Cupit, P.M., Wallance, Y.P., 1997. Emerging production system for antibody therapeutics. In: Harris, W.Q.J., Adair, J.R. (Eds.), Antibody Therapeutics. CRC Press Co.

Poul, M.A., Cerutti, M., Chaabihi, H., Devauchelle, G., Kaczorek, M., Lefranc, M.P., 1995a. Design of cassette baculovirus vectors for the production of therapeutic antibodies in insect cells. Immunotechnology 1, 18.

Poul, M.-A., Cerutti, M., Chaabihi, H., Devauchelle, G., Kaczorek, M., Lefranc, M.P., 1995b. Design of cassette baculovirus vector for the production therapeutic antibodies in insect cells. Immunotechnology 1, 189.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. A Laboratory Manual, 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savary, A.C., Georges, B., Auriault, C., 1999. Adaptation of BTI-TN5B1-4 (high five) insect cells for large-scale production in a stirred bioreactor. Biotechniques 27, 448.

Tan, W., Lam, P.H., 1999. Expression and purification of a secreted functional mouse/human chimaeric antibody against bacterial endotoxin in baculovirus-infected insect cells. Biotechnol. Appl. Biochem. 30, 59.

Towbin, H., Staehelin, T., Gordon, I., 1980. Electrophoretic transfer of protein from polyacrylamid gels to nitrocellulose sheets: procedures and some applications. Proc. Natl. Acad. Sci. USA 76, 4350.

Xiang, J., Robder, J., Pan, Z., Roifman, C., Hozumi, N., 1991. Modification in framework region I results in a decreased affinity of chimeric anti-TAG 72 antibody. Mol. Immunol. 28, 141.

Article No. jmbi.1999.3444 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 296, 57–86

JMB

Fully Synthetic Human Combinatorial Antibody Libraries (HuCAL) Based on Modular Consensus Frameworks and CDRs Randomized with Trinucleotides

Achim Knappik[1]*, Liming Ge[1], Annemarie Honegger[2], Peter Pack[1] Melanie Fischer[1], Günter Wellnhofer[1], Adolf Hoess[1], Joachim Wölle[1] Andreas Plückthun[2] and Bernhard Virnekäs[1]

[1]*MorphoSys AG, Lena-Christ-Str. 48, 82152, Martinsried/ Munich, Germany*

[2]*Biochemisches Institut Universität Zürich Winterthurerstrasse 190 CH-8057, Zürich Switzerland*

**Corresponding author*

By analyzing the human antibody repertoire in terms of structure, amino acid sequence diversity and germline usage, we found that seven V_H and seven V_L (four Vκ and three Vλ) germline families cover more than 95% of the human antibody diversity used. A consensus sequence was derived for each family and optimized for expression in *Escherichia coli*. In order to make all six complementarity determining regions (CDRs) accessible for diversification, the synthetic genes were designed to be modular and mutually compatible by introducing unique restriction endonuclease sites flanking the CDRs. Molecular modeling verified that all canonical classes were present. We could show that all master genes are expressed as soluble proteins in the periplasm of *E. coli*. A first set of antibody phage display libraries totalling 2×10^9 members was created after cloning the genes in all 49 combinations into a phagemid vector, itself devoid of the restriction sites in question. Diversity was created by replacing the V_H and V_L CDR3 regions of the master genes by CDR3 library cassettes, generated from mixed trinucleotides and biased towards natural human antibody CDR3 sequences. The sequencing of 257 members of the unselected libraries indicated that the frequency of correct and thus potentially functional sequences was 61%. Selection experiments against many antigens yielded a diverse set of binders with high affinities. Due to the modular design of all master genes, either single binders or even pools of binders can now be rapidly optimized without knowledge of the particular sequence, using pre-built CDR cassette libraries. The small number of 49 master genes will allow future improvements to be incorporated quickly, and the separation of the frameworks may help in analyzing why nature has evolved these distinct subfamilies of antibody germline genes.

Keywords: human antibodies; phage display library; combinatorial immunoglobulin repertoire; antibody expression; trinucleotide mutagenesis

Introduction

The selection of antibody fragments from libraries using enrichment technologies such as phage-display (Smith & Scott, 1993), ribosome display (Hanes & Plückthun, 1997), bacterial display (Georgiou *et al.*, 1997) or yeast display (Kieke *et al.*, 1997) has proven to be a successful alternative to classical hybridoma technology (for recent reviews,

Present addresses: L. Ge, Xerion Pharmaceuticals, Fraunhoferstr. 9, 82152 Martinsried, Germany; P. Pack, MTM Laboratories AG, Heidelberg, Germany.

E-mail address of the corresponding author: knappik@morphosys.de

0022-2836/00/60057–30 $35.00/0

EXP

see Winter *et al.*, 1994; Hoogenboom *et al.*, 1998; Spada *et al.*, 1997; Rodi & Makowski, 1999). Phage display was developed first (Smith, 1985) and has been improved the furthest, especially in the antibody field. It is likely that conventional hybridoma technology may be superceded by a combination of these technologies, since these approaches are faster, involve no animals, yield antibodies of at least comparable affinities and work also with self-antigens or toxic molecules (Hoogenboom *et al.*, 1998). The selection of antibodies must start from an initial, highly diverse library. Here, we describe the construction of such a library by total gene synthesis, based on a structural analysis of the human antibody repertoire.

Human antibodies are of particular interest, since they are considered to be valuable for therapeutic applications (Carter & Merchant, 1997), avoiding the HAMA (human anti-mouse antibody) response frequently observed with rodent antibodies. Although it has been demonstrated in many examples (Dall'Acqua & Carter, 1998) that chimerization or humanization of rodent antibodies through protein engineering can successfully retain the affinity and specificity of the parental molecule (Baca *et al.*, 1997), this strategy is time-consuming and still does not yield fully human antibodies.

Previous phage-display libraries of human antibodies have been generated from immunized donors (Barbas & Burton, 1996), germline sequences (Griffiths *et al.*, 1994) or, most recently, naive B-cell Ig repertoires (Vaughan *et al.*, 1996; Sheets *et al.*, 1998; De Haard *et al.*, 1999). Selection from these libraries by phage-display has yielded human antibodies against numerous haptens, peptides and proteins. While these libraries have all been successful, their uncontrollable composition and problems with the subsequent expression of the antibodies (see below) and restricted engineering possibilities made it desirable to use a complete protein engineering approach to solve the problem.

The success of obtaining high-affinity antibodies is generally assumed to be related to the initial library size (Perelson, 1989), even though the exact relation may not be tractable by theoretical considerations, as it may be antigen-dependent. Consequently, successful "one-pot" libraries have all been large (Griffiths *et al.*, 1994; Vaughan *et al.*, 1996; Sheets *et al.*, 1998; De Haard *et al.*, 1999). It is important to note that, obviously, only the functional library size, i.e. the number of correctly assembled clones without any frameshift, stop codon or deletion, will contribute to the diversity. This number can be orders of magnitude below the apparent diversity usually reported, which is normally obtained by counting the numbers of transformants.

It has been shown that the *Escherichia coli* expression yields of functional antibody fragments can vary dramatically, even if the antibody gene is expressed in the same format, vector and expression strain. This effect has been shown to depend on cellular folding, which in turn is influenced by the antibody sequence and can be successfully improved by protein engineering (Knappik & Plückthun, 1995). There is growing evidence that critical amino acid residues located in turns at the surface or at the variable-constant (V-C) interface are responsible for the misfolding, aggregation or even toxic effects on the *E. coli* cells, hence leading to poor expression yields. Mutating those residues improved expression titers several-fold, without adversely affecting the binding properties (Deng *et al.*, 1994; Knappik & Plückthun, 1995; Ulrich *et al.*, 1995; Jung & Plückthun, 1997; Nieba *et al.*, 1997; Forsberg *et al.*, 1997). As phage display depends on correctly folded antibodies, there is some selection against poor folders (Deng *et al.*, 1994; Jackson *et al.*, 1995; Jung & Plückthun, 1997; Bothmann & Plückthun, 1998), and thus the functional library size will be decreased. However, the selection is clearly not stringent enough to secure that all molecules selected from a phage display library will have acceptable folding properties. Thus, to maintain diversity and secure reasonable expression properties of the selected molecules, it would be advantageous to create antibody libraries starting from well-expressed frameworks. While such approaches have been reported (Pini *et al.*, 1998; Jirholt *et al.*, 1998), only single frameworks have been used in these attempts, and consequently, the structural diversity does not approach that of other naive libraries.

The humoral immune system, however, does not work by the "single-pot" approach (Nissim *et al.*, 1994), but rather uses an evolutionary strategy. The initial, antigen-independent variability is first generated during B-cell development by gene rearrangements (V(D)J-joining), leading to more than 10^9 different molecules at any one time in a human being (Winter, 1998). After a B-cell is activated, the antigen-driven process of somatic mutation is initiated (Rajewsky, 1996), and remarkable improvements in binding can be found. It has been shown that mutations occurring in CDRs 1 and 2 are preferentially selected (Wagner & Neuberger, 1996; Ignatovich *et al.*, 1997; Green *et al.*, 1998), as their diversity in the initial germline variants is much more limited than that of the CDR3s (Tomlinson *et al.*, 1996). The design of an artificial library should make it convenient to follow this same approach. Indeed, previous experiments with peptides (Cwirla *et al.*, 1997), RNA-aptamers (He *et al.*, 1996) and antibodies (Schier *et al.*, 1996a; Hanes *et al.*, 1998) have shown that the evolutionary approach and, in the case of antibodies, CDR walking (Yang *et al.*, 1995; Schier *et al.*, 1996a; Wu *et al.*, 1998) can dramatically improve affinities. However, in the absence of suitably engineered genes, such an optimization can be extremely laborious.

The human antibody germline repertoire has recently been completely sequenced. There are about 50 functional V_H germline genes located on chromosome 14 (Tomlinson *et al.*, 1992; Matsuda

& Honjo, 1996), which can be grouped into six subfamilies according to sequence homology. About 40 functional V_L kappa genes comprising seven subfamilies are located on chromosome 2 (Cox *et al.*, 1994; Barbie & Lefranc, 1998), and about 30 functional V_L lambda genes grouped into ten subfamilies can be found on chromosome 22 (Williams *et al.*, 1996; Kawasaki *et al.*, 1997; Pallares *et al.*, 1998). The groups vary in size from one member (e.g. V_H6 and Vκ4) to up to 22 members (V_H3), and the members of each group share a high degree of sequence homology. By comparing rearranged sequences of human antibodies with their germline counterparts we (this work) and others (Cox *et al.*, 1994; Ignatovich *et al.*, 1997) have found that many human germline genes are never or only very rarely used during an immune response.

In structural terms, the V_H and V_L domains comprising the antigen binding Fv moiety (see Figure 1) share a common fold that, in its central portions, is almost perfectly superimposable, even when fragments from different species are compared (Chothia *et al.*, 1998). Larger differences are observed only in the conformation of the CDRs, and it has been shown in a series of studies (Chothia & Lesk, 1987; Chothia *et al.*, 1989; Al-Lazikani *et al.*, 1997) that all CDRs except V_H CDR3 adopt only a few distinct conformations. Hence the repertoire of conformations is limited to a relatively small number of discrete structural classes, depending on both the CDR length and the so-called canonical amino acid residues (Chothia & Lesk, 1987).

Here, we report the design, construction and analysis of a novel human antibody library concept designated HuCAL (Human Combinatorial Antibody Libraries). Each of the human V_H and V_L subfamilies that is frequently used during an immune response is represented by one consensus framework, resulting in seven HuCAL master genes for heavy chains and seven for light chains, and thus 49 combinations. All genes were made by total synthesis, thereby taking into consideration codon usage, unfavorable residues that promote protein aggregation as well as unique and general restriction sites flanking all CDRs, leading to modular genes that contain readily accessible CDRs and can be easily converted into different antibody formats.

A first set of antibody libraries based on the HuCAL concept was created by randomizing both the V_H and V_L CDR3 encoding regions of the 49 master genes using trinucleotide cassette mutagenesis (Virnekäs *et al.*, 1994), which leads to high-quality libraries. The cassettes were designed such that the naturally occuring diversity was covered, both in terms of length and amino acid composition. The final HuCAL antibody libraries (HuCAL version 1) were extensively characterized by sequencing, expression behavior and numerous selection experiments against a wide variety of antigens.

Results

Analysis of the human antibody repertoire

Sequence analysis

Amino acid sequences from variable domains of human immunoglobulins were collected from Kabat (Kabat *et al.*, 1991; Johnson *et al.*, 1996;†) and Genbank (Benson *et al.*, 1997) and incorporated into three databases, V heavy chain (V_H), V kappa (Vκ) and V lambda (Vλ), and aligned, using the Kabat numbering system. For each of the three chain types, rearranged sequences were collected whenever more than 70 positions had been determined, giving 386, 149 and 675 entries for Vκ, Vλ and V_H, respectively, at the time of library design. Similarly, all germline sequences were collected (48, 26 and 43 entries for Vκ, Vλ and V_H, respectively), as the complete loci‡ (see Cook & Tomlinson, 1995), had not been published at that time. Finally, all known D and J sequences were collected. Although the design was started before the complete germline repertoire was known, the availability of the whole repertoire and a larger number of rearranged sequences would not have influenced the library design, which was demonstrated by repeating the analysis using the complete germline repertoire and a larger database (846, 413 and 1201 entries for Vκ, Vλ and V_H, respectively) of human rearranged sequences (see Figure 2).

The binning into families is somewhat arbitrary, depending on how the homology cutoff between families is defined. Initially, for Vκ, seven families were established, Vλ was divided into eight families and V_H into six families. The single V_H germline gene of the V_H7 family (van Dijk *et al.*, 1993) was included in the V_H1 family, since the genes of the two families are highly homologous. Upon more detailed analysis, regarding canonical CDR conformations and canonical framework residues as well as gene usage (see below), the number of families was raised to seven for V_H, but was reduced to four for Vκ and three for Vλ.

To further examine the concept of constructing HuCAL using the equidistant partitioning of sequence space as an efficient means to engineer library diversity, it was important to test the usage of the structural groups in actual rearranged genes of antibodies. By counting the number of differences between each rearranged entry and each germline sequence, the nearest germline counterpart was identified for each rearranged sequence. Altogether, 532 (79 %) V_H sequences and 474 (86 %) V_L sequences (343 Vκ and 131 Vλ) could be clearly assigned to germline counterparts.

† http://immuno.bme.nwu.edu/

‡ Now available at VBASE, http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html

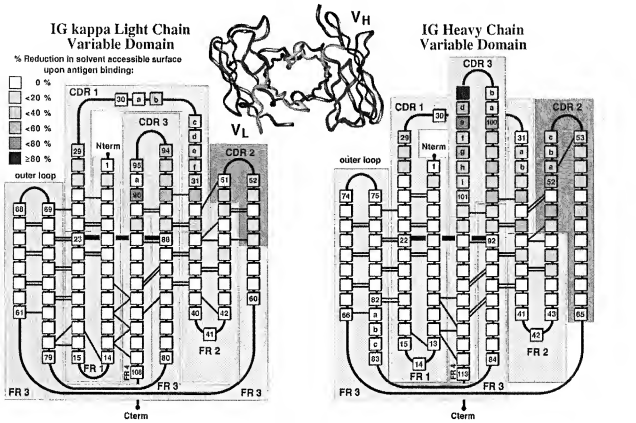


Figure 1. A representation of V_L and V_H structures, consensus hydrogen bonding pattern and antigen contacts. Residues are color-coded (from white to red) to indicate the reduction of residue solvent-accessible surface upon antigen binding, averaged over 52 liganded structures taken from the Brookhaven protein structure database (http://www.rcsb.org/pdb/). Residue numbering is according to Chothia *et al.* (1992), Tomlinson *et al.* (1995) and Williams *et al.* (1996), and CDR definitions conform to Kabat *et al.* (1991). Complementarity determining regions CDR1, CDR2 and CDR3 are indicated by blue, green and pink coloring, and framework regions by gray underlays.

Our results (see Table 1 and Figure 2) confirm the biased usage of human germline genes analysed previously (Tomlinson *et al.*, 1992; Cox *et al.*, 1994; Ignatovich *et al.*, 1997). The V_H germline gene usage was found to be restricted to about 12 genes from five sub-families, which are used in approximately 80% of all cases. The V_H2 family is only rarely used. Only four of the Vκ germline families were found to be used, and out of these only seven genes were used frequently (81%). The Vλ germline gene usage was found to be restricted to three families, which are used in 93% of all cases, and five genes from these three families were used most frequently (Table 1). We concluded that the vast majority (98% of all V_H, more than 99% of all Vκ and more than 93% of all Vλ) of human antibodies are derived from only five V_H and seven V_L families (four Vκ and three Vλ). Although the three germline genes of the V_H2 family are not frequently used, we decided to cover all six V_H families with our consensus approach, and therefore we included this family for further analysis.

The strategy of the synthetic library approach was therefore to represent each family by one representative member, subject to verification of the structural consequence of the distribution of CDR conformations (see the next section).

Structural analysis

Despite their great variability in length and sequence, the conformation of the antigen binding loops, denoted CDR (complementarity determining regions), have been shown to adopt only a limited number of main-chain conformations, termed canonical structures (Chothia *et al.*, 1989). The adopted structure depends on both the CDR length and the identity of certain key amino acid residues, both in the CDR and in the contacting framework, involved in its packing. The six V_H, four Vκ and three Vλ germline families, as defined above from the dendrogram analysis, were therefore analyzed for the canonical structures of CDRs that they were predicted to encode, in order to define the structural repertoire covered by these families (Table 1). In

Table 1. Frequency of germline family usage and corresponding types of canonical structures

		Frequently used germline genes			Canonical structure prediction		Chosen HuCAL canonical structures	
Subfamily	Family usage (%)	Locus	DP name	Usage (%)	CDR1	CDR2	CDR1	CDR2
		1-69	DP-10	6		H2-2		H2-2
VH1	19	1-18	DP-14	4	H1-1	H2-3	H1-1	H2-3
		1-02	DP-8	4				
VH2	2	-	-	-	H1-3	H2-1	H1-3	H2-1
		3-23	DP-47	12		H2-1		
VH3	34	3-30.3	DP-46	5	H1-1	H2-3	H1-1	H2-3
		3-48	DP-51	3		H2-4		
		4-34	DP-63	5	H1-1			
VH4	12				H1-2	H2-1	H1-1	H2-1
		4-59	DP-71	4	H1-3			
		5-51	DP-73	16				
VH5	19				H1-1	H2-2	H1-1	H2-2
		5-a	-	3				
VH6	14	6-01	DP-74	14	H1-3	H2-5	H1-3	H2-5
		O12	DPK9	9	L1-2	L2-1	L1-2	L2-1
Vκ1	32							
		O8	DPK1	7				
Vκ2	7	A3	DPK15	4	L1-3	L2-1	L1-4	L2-1
					L1-4			
		A27	DPK22	29	L1-2			
Vκ3	51	L6	-	11	L1-6	L2-1	L1-6	L2-1
		L2	DPK21	10				
Vκ4	10	B3	DPK24	10	L1-3	L2-1	L1-3	L2-1
Vκ5-7	0	-	-	-	L1-2	L2-1	-	-
		1b	DPL5	13	13			
Vλ1	31				14	7	13	7
		1c	DPL2	11				
		2a2	DPL11	18				
Vλ2	33				14	7	14	7
		2e	DPL12	11				
Vλ3	29	3r	DPL23	15	11	7	11	7
					12	7		
Vλ4-10	8	-	-	-	13	11	-	-
					14	12		

The human immunoglobulin germline subfamilies are listed together with their percentage usage as calculated by comparison with rearranged sequences. The percentage usage is determined from using the initial database of rearranged sequences with 1006 entries. The percentage usage calculated from the updated database with 2460 entries is given in Figure 2. The most frequently used germline genes according to our analysis are also given (locus name as well as DP nomenclature, see Tomlinson *et al.* (1992) for V_H, Cox *et al.* (1994) for Vκ, and Williams *et al.* (1996) for Vλ) together with their corresponding usage (derived from analysis of the smaller database). For details of the calculation, see the text. The canonical conformations that are present in each subfamily are shown together with the canonical conformations that have been chosen for HuCAL design. The canonical structure nomenclature is according to Chothia *et al.* (1992) for V_H, Tomlinson *et al.* (1995) for Vκ, and Williams *et al.* (1996) for Vλ.

the following, we will use the CDR definitions given by Kabat *et al.* (1991) (see also Figure 1) and the sequence numbering according to structural criteria defined by Chothia (Chothia *et al.*, 1992; Tomlinson *et al.*, 1995; Williams *et al.*, 1996).

The structural repertoire of the human V_H sequences was previously analyzed in detail by Chothia *et al.* (1992). In total, three conformations of CDR1 (H1-1, H1-2 and H1-3) and five conformations of CDR2 (H2-1, H2-2, H2-3, H2-4 and H2-5) have been defined, and the observed combinations have led to the conclusion that almost all sequences have one of seven main-chain folds. For the highly diverse CDR3, which is encoded by the D and J-minigene segments and uncoded nucleotides (N-region diversity), structural families have been defined only very recently (Morea *et al.*, 1998; Oliva *et al.*, 1998), but structural predictions are not approaching the accuracy seen for the canonical folds of the other CDRs.

All members of the V_H1 family encode the CDR1 conformation H1-1, but differ in their CDR2 conformation: both the H2-2 and the H2-3 conformation were found in five germline genes. Since these two types of CDR2 conformations are defined by different types of amino acids at position 71 located in framework 3, we divided the V_H1 sub-family into two further sub-families: V_H1A with CDR2 conformation H2-2 (alanine at position 71) and V_H1B with the conformation H2-3 (arginine at position 71). Upon model building (see below), we decided to include both gene types into the library design and to construct both a V_H1A and V_H1B master gene (see below).

The members of the V_H2 family were all predicted to have the conformations H1-3 and H2-1 in CDR1 and CDR2, respectively.

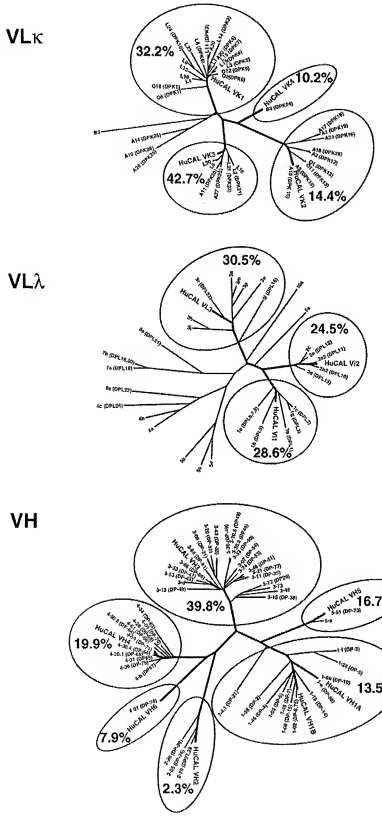


Figure 2. Coverage of germline sequence space by HuCAL sequences. The protein sequences representing the human V_L and V_H germlines were taken from VBase (http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html) and aligned to the 14 HuCAL sequences. The Phylip (http://evolution.genetics.washington.edu/phylip.html) and ClustalW (see ftp://ftp.ebi.ac.uk/pub/software/mac/clustalw/) phylogeny program packages were used to generate separate unrooted trees for the V_L kappa, V_L lambda and V_H sequences. Percentages indicate the fraction of rearranged sequences in the database that cluster within the different germline subgroups. For these calculations, we used a database of rearranged sequences with 846 V_L kappa, 413 V_L lambda and 1201 V_H sequence entries. The difference to 100% in the case of V_L kappa (0.5%) and lambda (16.4%) is due to rarely used germline subfamilies that are not represented by the HuCAL master genes.

The CDR1 conformation of the V_H3 family members was predicted in all cases to be H1-1, but three different types were found for CDR2 (H2-1, H2-3, H2-4). In these CDR2 conformations, the canonical framework residue 71 is always arginine, while the loop conformation of CDR2 is defined by

the residues 52a and 55 as well as the length variation. Of the rearranged V_H3 sequences, 80% were predicted to contain the H2-3 conformation. Therefore, the V_H3 family is best represented by a sequence containing the canonical conformations H1-1 and H2-3, even though the more groove-like shapes of binding sites from the longer CDR-H2 types may be introduced later by CDR shuffling.

The V_H4 family members were predicted to contain three types of CDR1 conformations; namely, H1-1, H1-2 and H1-3. The CDR1 canonical framework residue 26 was found to be glycine in all cases, and the CDR1 loop conformation is defined solely by residues located in this region. Since 62% of all rearranged V_H4 sequences contained the H1-1 type of CDR1, this conformation was chosen for representing the V_H4 family. The CDR2 conformation of the V_H4 members was found to be H2-1 in all cases.

The two members of the V_H5 family were found to have the conformation H1-1 and H2-2, and the single germline gene of the V_H6 family had the conformation H1-3 and H2-5 in CDR1 and CDR2, respectively. Hence, in structural terms the majority of the frequently used members of the six V_H families can be represented by seven sequences, since only the V_H1 family contained two types of canonical CDR folds defined by residues in the framework region, and since V_H3 and V_H4 were decided to be represented by the most prevalent type. The canonical conformations not present in the design can be incorporated later during CDR library generation, since the key residues for those conformations are part of the CDR itself.

The structural repertoire of the human Vκ germline sequences was analyzed by Tomlinson *et al.* (1995). There are four conformations of the CDR1, which are defined by the length of the loop (7, 8, 12 and 13 amino acid residues) and the nature of residues 2, 25, 29, 33 and 71. The CDR2 loop of human Vκ domains is only three amino acid residues long in all cases, and is predicted to adopt a single canonical fold. Most human Vκ germline segments encode also a single conformation of the CDR3 loop, which is stabilized by the conserved *cis*-proline 95, but other conformations in rearranged sequences are possible due to the process of V-J joining and the potential loss of this proline residue. Since the CDR3 region was planned to be randomized for library generation, this area was not considered for the consensus sequence design. Hence, the structural repertoire of Vκ domains is essentially defined by the conformation of the CDR1 region. All members of the Vκ1 family contained a seven residue CDR1 (L1-2), and the most frequently used members of the Vκ2 family contained a 12 residue CDR1 (L1-4). The members of the Vκ3 family contained either a seven (L1-2) or an eight (L1-6) residue CDR1. Since the canonical framework residues that additionally define the CDR1 conformation are identical in both cases, and since more than 60% of the rearranged Vκ3 sequences contained the CDR1 conformation L1-6, this type was chosen for the consensus sequence. The single germline member of the Vκ4 family contained a 13 residue CDR1 (L1-3).

The structural repertoire of the human Vλ germline sequences was analyzed by Williams *et al.* (1996). The three families analyzed here encode identical conformations of the CDR2 loop. The CDR3 loop conformation is thought to be more highly variable, as there is some length variation and no *cis*-proline residue. Since this part was planned to be randomized for library generation, this area was not considered for the consensus sequence design. Although the CDR1 region of the Vλ1 family contains either 13 or 14 amino acid residues, it is thought to adopt a single conformation, since the canonical key residues are conserved and the additional insertion of one residue has little effect on the overall structure (Chothia & Lesk, 1987). A CDR1 length of 13 residues, which was found in more than 90% of all rearranged Vλ1 sequences, was chosen for the Vλ1 consensus. The members of the Vλ2 and Vλ3 families each encode a single defined type of CDR1 loop structure: the Vλ2 family encode a CDR1 loop of 14 residues, and the CDR1 loop length of the Vλ3 family is 11 residues.

In summary, from the eight different pairs of CDR1-CDR2 conformations encoded by the Vκ and Vλ germline genes that are used frequently, seven could be represented by four Vκ and three Vλ consensus genes. The remaining CDR1 conformation (seven residue CDR1 loop in the Vκ3 family) is not defined by canonical key residues in the framework region and can therefore be inserted into the Vκ3 consensus sequence during library generation. From the 11 different family-specific pairs of CDR1-CDR2 conformations found in the six V_H germline families, seven could be covered by dividing the family V_H1 into two families (V_H1A and V_H1B). The remaining four pairs (two in the V_H3 and two in the V_H4 family) were either not found in rearranged antibody sequences or are defined by the CDRs themselves and will therefore have to be created during the construction of CDR libraries. Hence, the structural repertoire of the human V genes used could be covered by 49 ($7\ V_H \times 7\ V_L$) different frameworks.

Design of consensus frameworks

The compilation of rearranged sequences was first divided into separate groups (four Vκ, three Vλ and seven V_H) according to the germline families described above. These protein sequence databases were used to compute the consensus sequences of each subgroup. By using the rearranged sequences instead of the germline sequences for calculating the consensus, the consensus was automatically weighted according to the frequency of usage. Additionally, frequently mutated and highly conserved positions could be identified.

For the CDR1 and CDR2 regions, the consensus of rearranged sequences was replaced with the amino acid sequence of one of the germline sequences of the corresponding family. This procedure removes any bias, as the CDRs of rearranged and mutated sequences are known to be mutated due to selection towards their particular antigens. In the case of Vλ, a few amino acid exchanges were introduced in some of the chosen germline CDRs in order to avoid structural constraints (position 30b in Vλ1 and positions 27 and 34 in Vλ3, see Figure 3).

To construct, assemble and verify the genes, as well as to obtain preliminary information on expression behavior, it was advantageous to first substitute the intended library of CDR3-H and CDR3-L cassettes with defined dummy sequences. We chose the sequences $_{89}$QQHYTTPP and $_{98}$WGGDGFYAMDY for the V_L and V_H chains, respectively, which are derived from the antibody 4D5 (Carter *et al.*, 1992a) and are known to be favorable for antibody folding in *E. coli* (Jung & Plückthun, 1997). Even though molecular modeling indicates that the omega loop from Vκ is not ideal in a Vλ framework because of steric clashes, good expression behavior could still be obtained, demonstrating the robustness of the frameworks (see below).

For the framework 4 regions, encoded by the J-elements, the consensus of the rearranged sequences in each family was calculated and found to be identical in all families of V_H and V_L (κ and λ). This shows that there is no correlation between V-usage and J-usage (Baskin *et al.*, 1998). In all three cases, this consensus sequence was identical with at least one of the naturally occurring sequences encoded by joining elements, indicating that the sequence is able to exist.

We have described, up to this point, only sequence information that was used to design the consensus sequences. It could therefore not be excluded that the consensus would lead to a molecule whose sequence might "jump" between different naturally occurring sequences, thereby creating certain artificial combinations of amino acid residues that are located far away in the sequence but give rise to contacts in the three-dimensional structure. It was therefore essential to verify the sequences by structural means. Otherwise, the uncritical use of the algebraic consensus might obscure a hidden interaction between certain residues, which can occur only in certain combinations. While this approach may also keep residues together that are linked only historically, it does safeguard against losing hidden long-range interactions (Saul & Poljak, 1993). As a first check, the most homologous rearranged sequence for each consensus sequence was identified by searching against the compilation of rearranged sequences, and all positions where the consensus differed from this nearest rearranged sequence were inspected (see Materials and Methods). Furthermore, models for the seven V_H and seven V_L consensus sequences were built and analyzed according to their structural properties (see the next section). As a result of this analysis, the following residues were exchanged (given is the position according to Kabat's numbering scheme, the substitution performed, and the name of the gene family): $S_{H65}T$ (V_H2), $N_{L34}A$ (Vκ1), $G_{L9}A$, $D_{L60}A$, $R_{L77}S$ (Vκ3) and $V_{L78}T$ (Vλ3).

After the consensus protein sequences were designed, phylogenetic trees were built with the programs PHYLIP† and ClustalW‡ (Thompson *et al.*, 1994). For this representation, we repeated the analysis of germline usage based on an updated database of rearranged human antibody sequences that was more than twice the size of the original database that we used for the design of the HuCAL sequences. Separate unrooted trees were built for the V_Lκ, V_Lλ and V_H sequences (Figure 2). This analysis illustrates the strategy adopted in the present study, which is an attempt to approach a more equidistant representation of sequence space, by having only one member for each of the main "branches" of the tree. By analyzing each consensus sequence as if it were a member of the germline, its position in the sequence map is indicated, and that it truly represents the family (Figure 2).

Molecular modeling and analysis

To obtain more information about the packing, CDR conformations and framework properties, all seven V_H frameworks, all four Vκ frameworks and the three Vλ frameworks were built *via* homology modeling. As a basis, a complete structural alignment of the approximately 100 independent antibody sequences available in the PDB (Bernstein *et al.*, 1977) was carried out as indicated in the legend to Figure 3. Usually, the template with the highest resolution and the fewest mutations relative to the consensus sequence to be modeled was used. For all models, multiple templates were compared, such that the effect of mutations in any of the templates could be evaluated directly from the structural alignment. The experimental structures displaying the highest degree of similarity to each of the HuCAL constructs are listed in Table 1 of the Supplementary Material.

In the models (see Figure 4), the dummy CDR3 sequences from the antibody hu4D5 (version 8) are shown (PDB file 1FVC). All models were checked with the program PROCHECK§ (Morris *et al.*, 1992; Laskowski *et al.*, 1993) and were shown to have no more residues in the less favorable regions of the Ramachandran plot than the template structures (some unfavorable torsion angles in loop regions

† see http://evolution.genetics.washington.edu/phylip.html

‡ ftp://ftp.ebi.ac.uk

§ http://www.biochem. ucl.ac.uk/ ~roman/procheck/procheck.html

are highly conserved, e.g. position 51 at the tip of CDR2 in V_L), as well as having no obvious cavity or unusual exposed hydrophobic region, and a full set of standard variable domain hydrogen bonds.

Consistent with sequence considerations, the great majority of canonical structures was predicted to be present by model building, when comparing the critical residues with the templates. More recent work (unpublished results), based on previous experimental observations from X-ray crystallography (Saul & Poljak, 1993) and mutagenesis (Langedijk *et al.*, 1998), has uncovered several more structural relationships within each V_H domain, which may to contribute to diversity. Particularly, relationships between the nature of the residues H6, H7 and H9, due to the different hydrogen bonding pattern of H6 to the backbone, can transmit a conformational change through the protein *via* residues H18, H82, H67, and H63. Our analysis showed that all types of conformations that occur commonly in natural human frameworks are represented in the chosen consensus frameworks.

In the V_H3 group of germline sequences, there is more variation in CDR2, because of the length variation of a two amino acid residue insertion occurring in a group of human sequences (positions 52b and c). These antibodies might form more cleft-like binding pockets, and this diversity is not present in the original library design, even though many other combinations of frameworks would be able to form cavities and clefts as well. Through the modular design, however, these longer CDR2 elements can easily be introduced by cassette mutagenesis.

An analysis in analogy to that reported by Nieba *et al.* (1997) showed that the exposed residues at the V/C interface are already of low hydrophobicity in all consensus frameworks, consistent with their superior expression behavior in *E. coli* (see below). Moreover, many of the residues identified as crucial for stability and clearly selectable by phage display, such as P_{L8} (defining a conserved kink in the first β-strand with a *cis*-peptide bond in Vκ domains, or the *trans*-proline residues at positions 8 and/or 9 in Vλ domains, see Spada *et al.*, 1998) are present in all master sequences. Residue R_{H66}, which is part of a conserved charge cluster, and frequently K in murine antibodies, where it leads to lower stability (see Proba *et al.*, 1998), is present in all master genes except V_H5, where the consensus was found to be Q_{H66}. All residues known to make conserved side-chain hydrogen bonds are present in the master genes. Side-chain to side-chain: R_{H38} to Q_{H46}; D_{H86} and Y_{H99}; R_{H66} to D_{H86}; R_{H94} to D_{H101}; Q_{L6} to T_{L102}; Q_{L37} (L_{L37} in Vκ2) to Y_{L86}; R_{L61} to D_{L82}. Side-chain to main-chain CO: R_{H66} to H_{82a}; T_{H87} to X_{H84}; Q_{L6} to X_{L86}; Q_{L38} to X_{L42}. Main-chain NH to side-chain: X_{H69} to Y_{H59}; X_{H75} to D_{H72}; X_{H83} to D_{H86}; X_{H92} and X_{H106} to E_{H6} or Q_{H6}; X_{H111} to T_{H87}; X_{L79} to D_{L82}; X_{L88} and X_{L101} to Q_{L6}. Interdomain: Q_{L38} to Q_{H39}. In this listing, X refers to positions without dominant residue preference.

The relative orientation of V_L with respect to V_H is still understood only poorly, and will depend on the exact pairwise combination and on the specific CDR3 sequences. Frequently, monoclonal antibodies are found with mutations within the interface. This introduces further uncertainty in building a model of the combining site, because a small deviation in angle can have a large effect at the top of the binding site. This variability of the relative orientation of the two domains is particularly large for Vλ domains and Vκ lacking the *cis*-Pro in position L95, and is further modulated by non-tyrosine residues in position L49. The "elbow" of ordinary Vκ CDR3 inserts around L96 into a notch in V_H and restricts the flexibility of the interface. Since the interface residues are highly conserved between all the consensus antibodies (see Figure 3), and since very similar frameworks are available as templates in the database, more reliable models may be possible for HuCAL antibodies than for antibodies further away from the consensus. This system of defined frameworks might, in addition, provide excellent access to studying this question of domain orientation experimentally.

Construction of the seven V_H and seven V_L master genes

The final result of the analysis described above was a collection of 14 amino acid sequences, which represent the frequently used antibody repertoire of the human immune system. These sequences were then back-translated into DNA sequences. In a first step, the back-translation was carried out using only codons that are known to be used frequently in *E. coli*. In a second step, these gene sequences were then examined for all possible restriction endonuclease sites, which could be introduced without changing the corresponding amino acid sequences. This was done by creating a database of all possible silent cleavage sites for each gene. Using this database, cleavage sites were selected that were located close to the CDR and framework borders and that could be introduced into all V_H, Vκ or Vλ genes simultaneously at the same position. This was considered essential to the overall strategy, as CDRs (or frameworks) can then be shuffled within pools of sequences, without even knowing the individual antibody sequence. In a few cases it was not possible to find a common cleavage site for all genes at one of the flanking regions. In that case, one amino acid residue of the sequence was changed if this change seemed to be feasible according to the available sequence and structural information as delineated in the molecular modeling section. Each sequence was then analyzed again after exchange as described above.

In total, six amino acid residues were exchanged during the design of the genes: $T_{H3}Q$ (V_H2), $S_{H82}G$ (V_H6), $E_{L1}D$ and $I_{L58}V$ (Vκ3), $K_{L24}R$ (Vκ4) and $T_{L22}S$ (Vλ3). Additionally, the first two amino acid residues of all three Vλ sequences were changed to

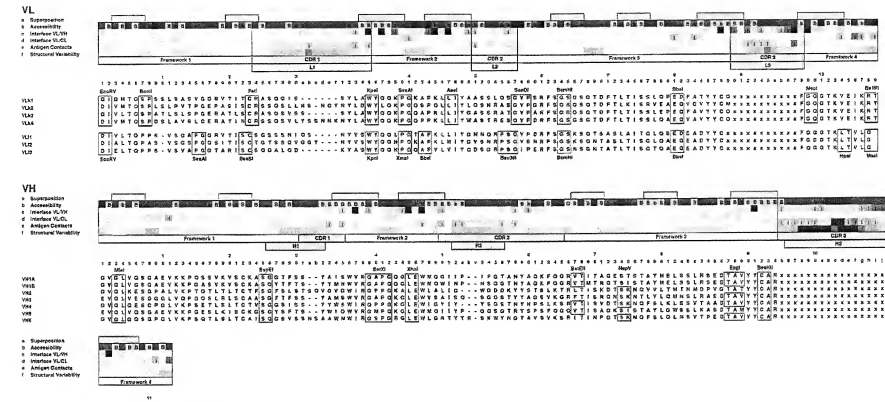


Figure 3. (*Legend opposite*)

aspartate-isoleucine in order to introduce an *Eco*RV site common to all V_L genes. After this design, only one element junction remained where no common cleavage site could be found. For this region (the border between CDR2 and framework 3 in the V_H sequences), two different types of cleavage sites were used instead: *Bst*EII for V_H1A, V_H1B, V_H4 and V_H5; and NspV for V_H2, V_H3, V_H4 and V_H6.

During this analysis, several potential restriction endonuclease sites were identified that could be introduced into every gene of a given group without changing the amino acid sequence, but which were not located at the flanking regions of the CDR or framework elements. The introduction of these cleavage sites made the system more flexible for further improvements. Finally, each gene sequence was modified again to remove, with the exception of the common restriction sites, all but one of the other sites (with a length of the recognition site of five or more bases), since this unique site might be used as a "fingerprint site" to differentiate the genes by restriction digest. All these changes were again carried out without changing the corresponding amino acid sequence. The 14 final protein sequences, including the introduced restriction pattern are shown in Figure 3.

The resulting consensus protein sequences were finally compared to the germline sequences, and a mean deviation of all 49 consensus sequences from their closest germline counterparts of 4.9(±3.6) residues was found. Thus, these consensus sequences are, on average, much more related to the germline sequences than the majority of rearranged sequences found in the database (mean deviation 14.7 amino acid residues). In contrast to the "original" germline sequences, however, our synthetic versions have all the advantages of sequences with known and predictable unique restriction sites at the framework/CDR borders.

The consensus gene fragments were then assembled from oligonucleotides by SOE-PCR assembly (see Materials and Methods for details). Gene segments encoding the human constant domains C_H1 (sub type IgG1), Cκ and Cλ1 were designed with optimized *E. coli* usage and synthesized in order to create F_{ab} fragments for display or expression (see Materials and Methods). After synthesis, the gene fragments were assembled and inserted individually into the expression vector pBS12, yielding 49 single-chain Fv genes containing identical dummy V_H and V_L CDR3s. The general format of the scFv genes is shown in Figure 5. All 49 master genes were also cloned in the reverse oriented scFv format (V_L-V_H) as well as in the F_{ab} format for future libraries (data not shown).

E. coli expression analysis

The *E. coli* expression of the 49 scFv genes (all containing the same V_H and V_L CDR3s from the antibody hu4D5, see Carter *et al.*, 1992a) was studied similarly as described by Knappik & Plückthun (1995). We found that all 49 master genes could be expressed as soluble proteins in the periplasm of *E. coli*, yielding a band of the correct size in FLAG Western blots of soluble *E. coli* crude extracts (data not shown). This indicates that all 49 combinations are most likely capable of forming V_H/V_L pairs, since unpaired domains tend to aggregate (Wall & Plückthun, 1999).

The ratio of soluble to insoluble expressed protein was quantified from Western blot experiments for each scFv gene, since this value has been shown to be correlated to the expression behavior

---

Figure 3. Protein sequences of the HuCAL V_H and V_L master genes. An alignment of the seven V_L and seven V_H sequences is shown, together with the approximate location of restriction endonuclease sites that were introduced into the corresponding DNA sequences. The alignment, numbering and loop regions (L1-L3, H1-H3) are according to structural criteria defined by Chothia *et al.* (1992), Tomlinson *et al.* (1995) and Williams *et al.* (1996). The H3 loop is given as defined by Chothia & Lesk (1987), although more recently, the extended H3 loop has been defined to include residues 92 and 104 (Morea *et al.*, 1998). CDRs are according to Kabat *et al.* (1991). Color codes indicate: (a) the structurally least variable regions used for least-squares superposition of the C coordinates of structures and models (residues L3-L7, L20-L24, L33-L39, L43-L49, L62-L66, L71-L75, L84-L90 and L97-L103 for V_L; H3-H7, H19-H23, H34 to H40, H44-H50, H67-H71, H78-H82, H88-H94 and H102-H108 for V_H) indicated as gray bars. (b) The average relative side-chain solvent-accessibility in the isolated domains, indicating the average side-chain solvent-accessibility for each position: 100% indicates a solvent-accessible surface of the same side-chain in the context of a poly(Ala) peptide in extended conformation. Strongly buried positions (less than 30% of the side-chain surface is solvent-accessible) are additionally marked by B, semi-buried positions (less than 50% of the side-chain surface is solvent accessible) are additionally marked by b. (c) The average loss of side-chain solvent-accessible surface upon formation of the V_L/V_H dimer interface, indicating residues directly contributing to the dimer interface. Positions strongly buried upon interface formation (more than 80% of the residual solvent-accessible surface buried in the interface) are additionally marked by I, and semi-buried positions (more than 40% of the residual solvent-accessible surface buried in the interface) are additionally marked by i. (d) The average loss of side-chain solvent-accessible surface upon formation of the VL/CL and VH/CH interface in the Fab fragment. (e) Average loss of side-chain solvent-accessible surface upon binding of the antigen. Positions strongly buried upon antigen binding (more than 80% of the residual solvent accessible surface buried in the interface) are additionally marked by I, and semi-buried positions (more than 40% of the residual solvent accessible surface buried in the interface) are additionally marked by i. (f) Average deviation of the C^α positions of all V_L or V_H structures, respectively, in the PDB database (http://www.rcsb.org/pdb/) from the average C^α positions.

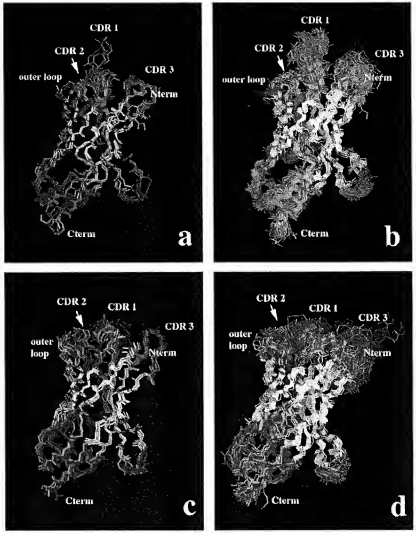


Figure 4. Coverage of the range of conformational variability of natural antibodies by the HuCAL frameworks. The homology models of the 14 HuCAL framework structures were generated using the program InsightII, modules Homology, Biopolymer and Discover (Biosym/MSI, San Diego, CA) as described in Materials and Methods. For CDR3, the sequence of antibody hu4D5 was used in all the models. The resulting V_L and V_H models were aligned by least-squares superposition of the C^{α} coordinates of residues L3-L7, L20-L24, L33-L39, L43-L49, L62-L66, L71-L75, L84-L90 and L97-L103 for V_L and H3-H7, H19-H23, H34 to H40, H44-H50, H67-H71, H78-H82, H88-H94 and H102-H108 for V_H (indicated in white). For comparison, 100 non-redundant V_L and V_H structures (mouse and human) were taken from the the RCSB protein structure database (http://www.rcsb.org/pdb/) and aligned. (a) HuCAL V_L models and (b) X-ray structures: cyan, kappa chains; blue, kappa chains lacking *cis*-Pro L8 (mouse only); pink, lambda chains. (c) HuCAL V_H models and (d) X-ray structures color-coded according to the sequence pattern correlating with the framework structure subtypes: magenta, H6 = Glu, H9 = Pro; pink, H6 = Glu, H9 = Gly; cyan, H6 = Gln, H9 = Ala; blue, H6 = Gln, H9 = Pro. The fourth conformation not covered by the HuCAL models shows some correlation with the presence of Pro in position H7, which is very rare in human sequences (<1%), but frequently seen in mouse sequences (in about 22% of the sequences).

of antibody fragments (Knappik & Plückthun, 1995; Nieba *et al.*, 1997; Jung & Plückthun, 1997). In each separate expression experiment, the HuCAL H3κ2 master gene was included as an internal control. The results are given in Table 2. The HuCAL genes were found to show a higher ratio of soluble to insoluble protein than many antibody genes obtained from natural monoclonal antibodies and subsequently expressed in *E. coli*. The ratio of soluble to insoluble protein ranges from 33% ($H_4\lambda 2$) to 90% ($H_1A\lambda 1$ and $H_6\kappa 1$), whereas a wide range of ratios has been found from natural antibody fragments, including many with ratios much below 30% under similar experimental conditions (Forsberg *et al.*, 1997; Nieba *et al.*, 1997). We could not find a correlation

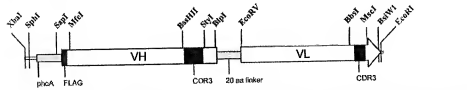


Figure 5. Arrangement of HuCAL scFv in the V_H-V_L orientation. The scFv gene cassette is preceded by a *phoA* signal sequence and a short FLAG tag. The two domains are fused by a 20 amino acid residue flexible linker. Some of the unique restriction sites common to all master genes are shown, and the location of the CDR3 regions is indicated.

between the type of V_L gene and expression behavior of the corresponding scFv genes, but it seemed that the genes encoding the V_H3 or V_H1A domains are showing higher soluble to insoluble ratios in almost all combinations (Table 2). These initial findings clearly need to be extended by a more detailed biophysical characterization.

The amounts of soluble protein produced, when compared to the H3κ2 gene set as 100 %, ranged from 26 % to 212 % (data not shown), indicating that soluble expression yields for all combinations fall into a narrow range.. Although we must expect that differences in the CDRs after randomization and selection of binders may influence the range of expression yields seen with the master genes, the use of well-expressed frameworks for creating libraries increases the chance to select well-expressed binding antibodies and reduces the large imbalances in the display efficiencies.

The CDR3 sequence introduced as dummy sequence in all V_L genes was taken from a V_L kappa gene (see above). Since this Vκ CDR3 contained a *cis*-proline residue at position 95, creating an omega-loop that is normally not found in V_L lambda CDR3s, and which might influence the folding and hence the expression behavior of the corresponding scFv genes, a Vλ dummy consensus CDR3 cassette encoding the sequence $_{89}$QSYDSSLS was designed and used to replace the Vκ dummy CDR3 in the H3λ1 scFv gene. Interestingly, however, no significant difference in expression yields could be detected (data not shown).

The expression behavior of two randomly chosen scFv genes (H2κ2 and H3κ2) was analyzed in more detail. These two genes were selected from panning experiments after library creation (see below) and therefore contained CDR3 sequences different from the dummy sequence of the master genes. Since both scFv fragments bound the antigen they were selected on, we could use ELISA experiments to determine the amount of active material in the lysates after different times of induction. The results are shown in Figure 6. We found that the expression titer after five hours of induction at 30 °C was 6 mg (H2κ2) and 10 mg (H3κ2) per liter of shaking-flask culture. The expression titer stayed constant for several hours and then decreased, probably due to the start of cell leakiness. This observed expression yield is significantly higher than that reported for antibody fragments from other libraries (Griffiths *et al.*, 1994; Vaughan *et al.*, 1996).

Design and construction of CDR3 library cassettes

Our rational approach to creating an antibody library aims at defining, with the smallest number of molecules possible, a structural diversity as large as possible. At the same time, it was important to design molecules that are likely to be stable and fold well. Furthermore, it was essential to direct the sequence diversity to those residues most likely in contact with the antigen. We decided for the first set of HuCAL libraries to randomize both CDR3 regions of the V_H and V_L genes simultaneously, since these two regions form the inner circle of the antigen binding site, and therefore show the highest frequency of antigen contacts in structurally known antibody-antigen complexes. In order to obtain the highest degree of diversity in

Table 2. Expression analysis of HuCAL master genes

	κ1	κ2	κ3	κ4	λ1	λ2	λ3
H1A	61	58	52	42	90	61	60
H1B	39	48	66	48	47	39	36
H2	47	57	46	49	37	36	45
H3	85	66 ± 6	76	61	80	71	83
H4	69	52	51	44	45	33	42
H5	49	49	46	67	54	46	47
H6	90	58	54	47	45	50	51

The amount of soluble full-length scFv relative to the total amount obtained is given (in %), as determined from quantitative Western blot analysis of HuCAL master genes expressed in *E. coli* after two hours of induction at 30 °C. The H3κ2 master gene which served as an internal control in each separate expression experiment was analyzed altogether 18 times. For this gene, the mean value and the standard deviation is given.

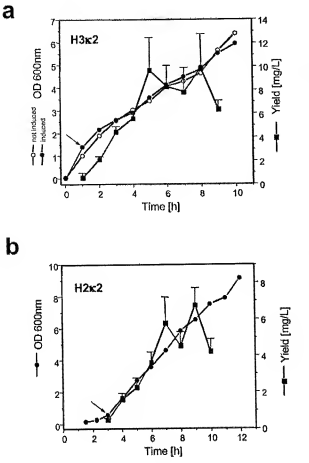


Figure 6. Growth curves and expression kinetics for two HuCAL scFv fragments. (a) Gene derived from the H3κ2 HuCAL framework; (b) gene derived from the H2κ2 HuCAL framework. The growth curves (circles) were determined by measuring the absorbance at 600 nm at the indicated time-points. For comparison, the growth curve of the uninduced culture (open circles) is given in (a). The arrows indicate the time-point of induction. The amount of functional scFv (squares) at the different time-points was determined by ELISA measurements of crude extracts. The corresponding purified antibody fragments of known concentration, also measured in the presence of a corresponding amount of cell extract, served as internal standard to calculate the scFv amount based on the ELISA signal obtained. The mean and standard deviation of three different measurements is given for each experiment.

the V_H CDR3, which is also the most variable region in natural antibodies, we applied the following strategy for library generation: first, we designed V_L CDR3 library cassettes strongly biased for the known natural distribution of amino acids (see below) with relatively low complexity and inserted those in the V_L master genes, aiming at a library size of about 10^7 members. Subsequently, we used these V_L libraries to insert a V_H CDR3 library cassette with very high complexity (both in terms of sequence composition and length variation), ensuring that every single library member contains a unique V_H CDR3 sequence.

Since we used trinucleotides (Virnekäs *et al.*, 1994) for the generation of the CDR3 library cassettes (see below), we could introduce any amino acid bias at any position of the cassettes. We decided to first analyze the sequence variability in the CDR3 regions of our databases of human rearranged antibody sequences and use this information together with structural data for the library design, in order to bias the CDR3 sequences towards the naturally found human antibodies.

Vκ CDR3

A total of 382 sequences of rearranged antibodies from our initial internal database were analyzed. In the following discussion, we will use the numbering system and definitions of CDRs regions introduced by Kabat even though this does not always correspond to the structural definitions (Chothia & Lesk, 1987; Barre *et al.*, 1994; Giudicelli *et al.*, 1997).

A fraction of 72.3 % of all CDR3s had a CDR length of eight amino acid residues, the remaining sequences had CDR lengths of less than seven (1.8 %), seven (7.3 %), nine (17.3 %), and ten (1.3 %) residues. Because of the predominance of CDRs of eight residues, we decided to consider just that size for constructing a CDR3 library. The omega-loop structure of Vκ CDR3 is determined by a characteristic *cis*-proline residue at position 95, which is encoded in 96 % of all κ germline genes, but can be lost upon V-J rearrangement. A total of six canonical structures have been discussed with structural data being available for structures 1 and 2 (Al-Lazikani *et al.*, 1997). In canonical structure 1, residues 90 and 95 are predominantly occupied by glutamine and proline, respectively, whereas in structure 2, the presence of *cis*-proline at position 94 is characteristic. About 87 % of all 382 sequences had Q_{L90}, and 78 % had P_{L95}, whereas P_{L94} was present in only 1 % of all sequences. Therefore, we decided to base the design of Vκ CDR3 on structure 1. Besides the canonical residues, position 89 showed a strong conservation, with glutamine present in 89 % of all sequences. Residues 89 and 90 are not part of the region outside the β-strand forming the CDR-L3 loop, which comprises residues 91 to 96 (Chothia & Lesk, 1987). Within CDR-L3, a high degree of variability (except for position 95 mentioned before) can be seen, with some preference for tyrosine at position 91. This corresponds well with the inspection of antigen contact residues in structurally known antibody-antigen complexes, showing that positions 91 to 94 and 96 seem to play the most important role (see Figure 3).

In our design of the library (see Figure 7(b)), we kept position Q_{L90} constant. Besides being a canonical residue, the side-chain of this glutamine residue does not contribute to the antigen-binding pocket, but points in the opposite direction. In the trinucleotide mixture, we biased positions 89 and

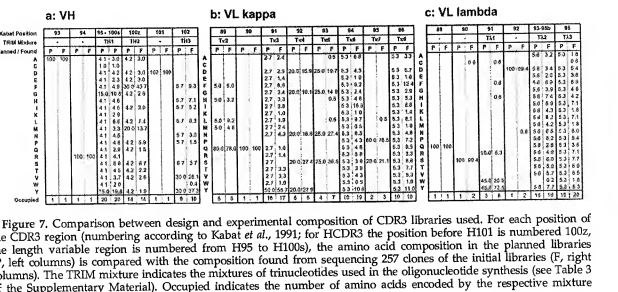


Figure 7. Comparison between design and experimental composition of CDR3 libraries used. For each position of the CDR3 region (numbering according to Kabat *et al.*, 1991; for HCDR3 the position before H101 is numbered 100z, the length variable region is numbered from H95 to H100s), the amino acid composition in the planned libraries (P, left columns) is compared with the composition found from sequencing 257 clones of the initial libraries (F, right columns). The TRIM mixture indicates the mixtures of trinucleotides used in the oligonucleotide synthesis (see Table 3 of the Supplementary Material). Occupied indicates the number of amino acids encoded by the respective mixture and found in the sequenced clones, respectively.

95 strongly towards glutamine and proline, respectively. A limited set of trinucleotide codons was allowed for positions 92 and 93, despite the fact that a large number of different residues can been found there, because the side-chains of these residues point away from the V_H CDR3 contact side. In contrast, for position 91, 18 amino acids (all except cysteine and proline) were allowed (biased towards Y_{L91}). Since proline is never found at position 91 in germline or rearranged sequences, it could be that P_{L91} would not allow the loop to form the correct conformation. Cysteine was omitted, since it was almost never found and it might cause problems during phage panning and later expression because of disulfide formation. Accordingly, for positions 94 and 96, all amino acids except cysteine were allowed. The residues located at those three most strongly randomized positions point into the binding pocket. By focusing the diversity towards positions that are most likely in contact with the antigen, we could reduce the overall theoretical diversity to a value of 1.3×10^6, which ensured that the theoretical diversity will be present in the final library.

For the four different HuCAL Vκ consensus genes, three trinucleotide-containing oligonucleotides were synthesized. A single oligonucleotide for Vκ1 and Vκ3 was possible, since both differ only at position 85 (κ1 T_{L85}; κ3 V_{L85}) and could thus be synthesized by using a mixture of two trinucleotides encoding threonine and valine in a 1:1 ratio at the appropriate position. Structural inspection revealed that residue 85 has no contact to other residues, thus making it likely that an exchange of these two similar amino acids would not cause problems. Indeed, we found the expected 1:1 ratio at this position after library construction and sequencing of clones (data not shown).

For oligonucleotide synthesis, six different trinucleotide mixtures (Tκ2 to Tκ6, see Figure 7(b)) had to be prepared comprising two to 19 codons, either biased or equally distributed. While initial results had suggested that different trinucleotides couple with different relative coupling yields (Virnekäs *et al.*, 1994), more controlled subsequent experimentation showed that these differences were not systematic (data not shown) and thus, trinucleotide mixtures were prepared directly using the desired molar ratios, thereby implicitely assuming an equal coupling yield. During oligonucleotide synthesis, the stepwise coupling ratio for trinucleotide mixtures ranged from 95.5% to 97.5%, the overall yield per oligonucleotide from 44% to 68%.

After cassette preparation, restriction digest and purification, the cassettes were ligated into the four Vκ consensus genes using the unique restriction sites *Bbs*I and *Msc*I, and the ligation mixtures were electroporated into *E. coli* TG1 cells. We obtained 6×10^6 independent colonies, and hence an almost complete coverage of the theoretical diversity. The quality of the cassettes was then checked by sequencing 235 independent clones. A total of 175 clones (75%) were completely correct and showed the library composition as planned. Four clones contained an unplanned amino acid at one position, which was most likely due to single-base mutations introduced during cassette preparation, three clones contained a one-base and six clones contained a one-codon deletion in the trinucleotide-encoded region. All other non-correct clones had the library cassette inserted twice or in the reverse orientation, or they contained one-base deletions in the 5′ mononucleotide region of the oligonucleotide. In order to obtain more statistical data on codon incorporation, all codons originating from trinucleotide positions were analyzed. Figure 7(b) shows the result of that analysis. Over-

all, the data are in excellent agreement with the expected distribution. Q_{L89}, Y_{L91}, and P_{L95} appeared almost exactly as planned at these strongly biased positions.

Vλ CDR3

A total of 147 rearranged human Vλ sequences were collected and analyzed. The lengths of the CDR3s (positions 89 to 96) ranged from seven to 12 residues, the majority (92%) having between eight and ten residues (L3 loop lengths six to eight according to Chothia & Lesk, 1987). Therefore, we decided to construct a CDR library comprising these three different length variants. Analysis of the amino acid composition in the rearranged sequences revealed a high degree of variability at positions 93 to 96, and to a smaller extent at positions 89 to 92. The inspection of antigen contact residues in the case of an antibody of canonical structure 1 (Chothia & Lesk, 1987, see Figure 3) revealed that positions 91, 94, and 96 seem to play the most important role. A single Vλ CDR3 oligonucleotide for all three Vλ consensus genes was designed, where Q_{L89} and S_{L90} were kept constant, since neither position is part of the loop region. Similarly, D_{L92} was fixed as the most frequent amino acid at that position and because its side-chain points away from the binding pocket. Residue 91, which packs against V_H CDR3, was limited to the three most frequent amino acids found in the database (arginine, tryptophan and tyrosine). At positions 93 to 95B, an equimolar mixture of all amino acids except for cysteine and tryptophan was allowed, since cysteine and tryptophan were never found in the rearranged sequences. Position 96 was completely randomized, except for cysteine.

Since the framework 3 region adjacent to the CDR3 of all three Vλ master genes is identical, we could use a single oligonucleotide for all three genes. For oligonucleotide synthesis, three different trinucleotide mixtures had to be prepared comprising three biased codons, 18 or 19 codons (in both cases equally distributed). The three mixtures and their positions in the CDR3 are given in Figure 7(c). On average, the stepwise coupling ratio for trinucleotide mixtures was about 98.9%, the overall yield for the oligonucleotide was 80%. During oligonucleotide synthesis, we used four consecutive sub-stoichiometric coupling steps at the triplet position corresponding to residue 95A. Thereby, we created an oligonucleotide with variable length covering CDR3 lengths between eight and 11 amino acid residues, with the smallest fraction having a CDR3 length of 11 residues. The theoretical diversity of these length variants ranged from 3.3×10^5 (eight residues) to 1.9×10^9 (11 residues).

After cassette preparation, restriction digest and purification, the cassette was ligated into the three Vλ consensus genes using the unique restriction sites *Bbs*I and *Hpa*I, and the ligation mixtures were electroporated into *E. coli* TG1 cells. We obtained 5.7×10^6 independent colonies.

As described above for Vκ, the quality of the oligonucleotide was checked by sequencing (183 independent clones). Again, about 26% of incorrect sequences could be identified, with errors of the type similar to those found in the Vκ CDR3s. A total of 74% of all clones, however, had completely correct CDR cassettes. The amino acid composition was again in very good agreement with the desired distribution, except for Y_{L91}, which was over-represented at the expense of W_{L91} (see Figure 7(c)). The length distribution was also analyzed: we found that the majority contained a CDR3 length of eight (36%) or nine (42%) residues, the rest had a length of ten (21%) or 11 (2%) residues.

V_H CDR3

For the highly variable V_H CDR3s, all available rearranged sequences were grouped together, irrespective of the individual sub-families. A total of 572 sequences were analyzed. The analysis revealed that only position H101 is stongly biased (toward aspartate in 82% of all cases). This is in agreement with the findings that R_{H94} and D_{H101} form a highly conserved salt-bridge (Searle *et al.*, 1995), and that these two residues are critical for the "kinked base" (Shirai *et al.*, 1996) or "bulged torso" (Morea *et al.*, 1998) structure of the CDR3 loop. D_{H101} was therefore kept constant, although this limits the structural variability to only a subset of CDR-H3 conformations, as other structures are seen in antibodies devoid of the R_{H94}-D_{H101} salt-bridge.

Again, the observed variability corresponds well with the information obtained by inspection of antigen contact residues, showing that positions H95 to H100y seem to play the most important role, whereas H100z is involved to a lesser extent (see the legend to Figure 7 for HCDR3 position nomenclature). Position H102 was found not to be important for antibody/antigen interactions (see Figure 3).

When designing the library cassette, we decided to base the composition of the trinucleotide mixtures for all positions except for H100z and H102 on the overall amino acid composition of the natural heavy chain CDR3s. Positions H100z and H102 were analyzed separately. This resulted in three different codon mixtures, named TH1 (for H95 to H100y), TH2 (for H100z), and TH3 (for H102). The compositions of these mixtures are given in Figure 7(a).

Analysis of the length variability of CDR3 (positions 95 to 102) showed a range between four and 28 residues with a maximum at 13.0. Wu *et al.* (1993) found a mean length of 11.6 residues in their analysis of human antibody sequences. To be able to cover such a broad spectrum of length variants, two separate oligonucleotides were synthesized using the sub-stoichiometric coupling approach to create the shorter library CDR3Ha,

comprising five to 22 residues and the longer library CDR3Hb, comprising nine to 28 residues. Since the two length variants were kept separated during library construction (see below), their use might be adapted to the antigen in question. Moreover, by mixing these two libraries appropriately, it is possible to mimic the natural length diversity. The final yields for oligonucleotides CDR3Ha and CDR3Hb were 68 % and 74 %, respectively, and the sub-stoichiometric coupling rates varied between 35 % and 55 %. Based on these coupling rates, a theoretical length distribution for the two libraries CDR3Ha and CDR3Hb was calculated (see Figure 8).

After cassette preparation, restriction digest and purification, the cassettes were inserted into the scFv libraries already containing the randomized V_L CDR3s described above. We mixed all four Vκ and all three Vλ libraries before HCDR3 insertion, but we kept the V_H consensus genes separate (except V_H1A and V_H1B, which were also mixed). Hence, 24 separate libraries were created (V_H1 to V_H6, each either with four κ or three λ genes, and each either with the short or the long HCDR3 cassette). After electroporation into *E. coli* TG1 cells, we obtained altogether 2.1×10^9 independent colonies.

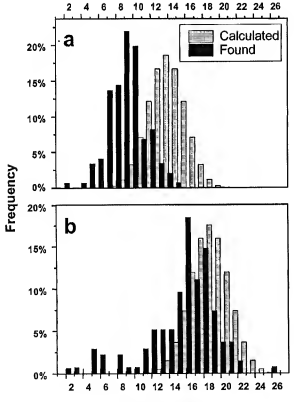


Figure 8. Distribution of CDR3 H length variants in the HuCAL1 libraries. The results from (a) the trinucleotide cassette HCDR3a, and from (b) the cassette HCR3b are shown (black columns) and compared to the length distribution as calculated from the substoichiometric coupling (gray columns). For details, see the text.

The quality of the V_H CDR3 s were checked by sequencing 257 clones. In Figure 7(a) the amino acid distributions for the trinucleotide mixtures TH1, TH2, and TH3 are given, showing again an excellent agreement with the calculated and designed frequencies. The sequencing results obtained from both V_H CDR3 length variants revealed that both library types follow a Gaussian length distribution, with the maxima at 9.0 and 16.6 residues (Figure 8). Thus, the actual length distribution was shifted towards shorter lengths when compared to the theoretical length distribution, but the whole range of naturally occurring length variants was covered by the two library variants.

The final library was designated HuCAL1. Altogether, we found the fraction of fully correct library members with CDRH3 and L3 as designed to be 61 %.

Diversity and binding constants

Phage-display as well as ribosome-display selection experiments were performed against a variety of antigens, including proteins, peptides, or whole cells. The HuCAL1 library comprosing all 49 combinations was used for selection experiments. Two or three panning rounds of phage display, or five or six rounds of ribosome display were performed in each case. After the final round, the selected scFv genes were subcloned as a pool in an expression vector and the transformants were screened for binding using ELISA or FACS assays. Details about the selection experiments and the characterization of binders will be given elsewhere (Krebs *et al.*, unpublished results; Hanes *et al.*, unpublished results). In the great majority of cases, many different scFv fragments could be identified, which bound the antigen specifically. The V_H and V_L framework usage for the first 250 specific binders selected from HuCAL1 *via* phage display is given in Table 3. All V_H and V_L frameworks could be selected, and so far 42 of the 49 framework combinations were found to be used. While the V_H4 gene segment is rarely used, the V_H3 gene segment predominates. The predominance of V_H3 occurs also in nature (see Table 1) and is even higher in other libraries (Griffiths *et al.*, 1994; Vaughan *et al.*, 1996). All other HuCAL frameworks seem to be used with similar frequency. There is also a considerable variation in V_H CDR3 length: the first 250 specific binders range from four to 24 residues (data not shown).

Selected binders were purified to homogeneity using affinity chromatography or IMAC, and their monovalent binding constants were measured using surface plasmon resonance (BIAcore). As shown in Table 4, binding constants of peptide binders were in the micromolar range, whereas affi-

Table 3. Framework usage

	κ1	κ2	κ3	κ4	λ1	λ2	λ3	Σ	%
H1A	4	1	3	3	1	3	7	22	9
H1B	3		4	2	1	6	1	17	7
H2	2	5	11	7	11	4	15	55	22
H3	12	5	16	11	15	7	25	91	36
H4					2		1	3	1
H5	4	1	5	3	1	5	14	33	13
H6	8		3	5	2	3	8	29	12
Σ	33	12	42	31	33	28	71	250	
%	13	5	17	12	13	11	28		

For each of the 49 HuCAL framework combinations, the number of specific scFvs from a collection of 250 binders against about 50 different antigens (haptens, peptides, proteins) is shown. All clones have been selected by phage display. The identity of the framework was determined by sequencing.

nities to protein antigens were usually in the low nanomolar range.

Discussion

Here, we describe the realization of the concept of fully synthetic human antibody libraries, designated HuCAL, which are built on seven V_H and seven V_L consensus frameworks, yielding 49 combinations in total.

We have extensively used these first libraries for the successful selection of highly specific binders against all kinds of antigens, including haptens, DNA, peptides, and proteins, including cell-bound receptor antigens (unpublished results). Intrinsic affinities down to the sub-nanomolar range were found against protein antigens, and the majority of binders were found to have dissociation constants between 1 and 1000 nM after only two rounds of selection. All frameworks have been selected, the selected antibodies could be shown to be expressed in good yields, they are suprisingly stable against thermal denaturation, and can be used in typical applications like ELISA, immunoblotting, FACS analysis, immunoprecipitation or immunohistochemistry even without any affinity maturation steps, verifying the successful design of completely synthetic human antibodies described in this study.

Strategy of modular design

The 49 consensus genes were derived by a stepwise analysis of human antibody sequences. First, the collected sequences were grouped into families according to sequence homology. Second, the

Table 4. Affinities of HuCAL scFvs

Antigen	scFv	Framework	Affinity (nM; BIAcore)	$k_{on} \times 10^5$ (M^{-1} s^{-1})	$k_{off} \times 10^{-2}$ (s^{-1})	App. size (kDa; SEC)
ICAM-1[a]	ICAM1-1	H3λ3	9.4	2.13	0.20	32
ICAM-1[a]	ICAM1-15	H5λ2	72.7	1.72	1.25	27
Insulin[b]	C59	H1Aκ1	0.082[e]	-	-	32
Insulin[b]	A21	H3κ2	8[e]	-	-	32
CD11b[c]	Mac1-5	H2λ1	1.0	7.92	0.09	36
CD11b[c]	Mac1-29	H2λ3	1.2	1.76	0.03	25
EGFR (human)	A9-1	H2κ2	246	1.34	3.30	25
Mac1 peptid[d]	3B2	H3λ2	1130	1.85	21.0	30
Hag peptide[d]	C22-2	H3κ4	610	1.41	8.6	28
NFκB peptide[d]	27HA1	H3λ3	1600	0.55	8.8	32

Affinity of FPLC-purified antibody monomers measured by SPR on a BIAcore biosensor. Antigens were coupled to CM5 sensor chips. In order to avoid contamination with multimeric variants, the monomeric scFv fragments were isolated by size-exclusion chromatography (SEC) (Krebs *et al.*, unpublished results; Hanes *et al.*, unpublished results).

[a] The extracellular part of human ICAM-1 (residues 26 to 479) was used as antigen.

[b] Selection against bovine insulin was carried out with ribosome display (Hanes & Plückthun, 1997; Hanes *et al.*, 1998; unpublished results). These antibodies carry additional point mutations created during PCR amplification.

[c] The I-domai of human CD11b (residues 149 to 353) was used as antigen.

[d] The following peptides were synthesized, coupled to a protein carrier and used for antibody selection.
Mac1 peptide: NH_2-C-DAPRSEKSRQELNTIASKPPRDHVF-COOH
Hag peptide: NH_2-C-AGPYDVPTDYASLRSHH-COOH
NFκB peptide: NH_2-C-LHVTKKKV-COOH

[e] Affinities determined with the inhibition BIAcore method, in which a mass transport-limited on-rate is measured as a function of antigen present in solution (Hanes *et al.*, 1998).

usage for each germline gene was analyzed by calculating for each rearranged sequence in the database the germline gene from which it was derived. Third, the families of frequently used antibody genes were analyzed in terms of structural diversity of the antigen binding loops, following the concept of canonical CDR conformations established by Chothia and co-workers (Chothia *et al.*, 1989). Fourth, consensus sequences were derived from the rearranged sequences, and grouped into families of frequently used human antibodies. Altogether, the analysis resulted in seven V_H, four Vκ and three Vλ consensus sequences, and our analysis suggests that this small set of consensus genes covers almost the entire structural repertoire encoded in those human antibody germline genes that were found to be used during the immune response.

Reducing the human antibody repertoire to 49 distinct Fv frameworks, yet without reducing structural diversity, made it feasible to obtain the sequences *de novo* by gene synthesis, which enabled us to incorporate several features into the genes that facilitate library construction, affinity maturation and *E. coli* gene expression. Moreover, the separate construction of the genes and the resulting libraries allowed detailed analysis of each master framework under defined conditions, which is not possible with antibody phage-display libraries derived from natural sequences by PCR cloning. Particularly, the presence of unique restriction sites across the whole library makes it possible to shuffle CDRs and frameworks, even at the level of pools, and without knowledge of the sequence of the antibodies. Furthermore, the approach is modular and can incorporate future knowledge of antibody structure, folding and stability, as individual framework pieces can easily be replaced in future versions.

The availability of separate libraries for each of the combinations allows one to analyze the performance of separate framework combinations and a direct comparison with results obtained from a mixture or the natural immune response. It also provides a way to force the selection against different epitopes on the same protein, which can be a very crucial feature given that *in vivo* applications may require the blocking of a binding site on a receptor by the antibody, while a different epitope on the receptor may be completely immuno-dominant. In this case, the preferentially selected but unwanted framework combination can simply be left out. Alternatively, separate affinity enrichments with subsets of frameworks can be carried out to enforce the binding of diverse epitopes. In addition, further analysis of the performance of this and other libraries may show that particular framework combinations contribute little to the pool of selected binders, while others need to be provided with more initial diversity in CDR1 and CDR2. The number of frameworks is, of course, arbitrary and can be adjusted by addition of new and subtraction of unnecessary ones.

Expression and folding properties

The HuCAL genes were adapted to *E. coli* codon usage. While we indeed found superior expression behavior from most of the synthetic genes, this probably reflects favorable protein folding properties (see below), although the avoidance of codons used only very rarely is at least a prerequisite for high expression yields. The consensus frameworks described here may be an interesting basis for elucidating the framework contributions to differences in folding yield during recombinant antibody expression and to thermodynamic stability. The absence of large differences in expression behavior between the consensus frameworks may improve library quality, since the probability of clones being eliminated during library selections due to very different effects of distinct antibody sequences on the bacterial cell physiology is minimized.

In this context, it is interesting to note that the high-expressing humanized antibody hu4D5, which was shown to be expressed 10-50-fold better in *E. coli* than the murine parental antibody (Carter *et al.*, 1992a), was designed using human consensus frameworks derived from the subfamilies V_H3 and Vκ1 (Carter *et al.*, 1992b). The human V_H3 germline gene 3-23 (DP-47), which is most homologous (99 % identity) to the HuCAL consensus amino acid sequence of the V_H3 germline subfamily, is also the most frequently used V_H3 germline gene (see Table 1) and it is very frequently found in antibody phage-display libraries based on human genes (Griffiths *et al.*, 1994; Vaughan *et al.*, 1996; Dorsam *et al.*, 1997; Boel *et al.*, 1998; Sheets *et al.*, 1998). Our theoretical analysis (unpublished results) showed that this framework has very few of the recognized sequence problems. Such problem spots include exposed hydrophobic residues that might promote misfolding and aggregation (Nieba *et al.*, 1997), non-Gly residues in positions with conserved positive phi angles, proline in position H40 (Knappik & Plückthun, 1995) and the disruption of the highly conserved charge cluster around R_{H66}/D_{H86} and R_{L61}/D_{L82} (Proba *et al.*, 1998).

It is reasonable therefore, to hypothesize that consensus sequences, which are closely related to phylogenetically old progenitor genes, are better adapted to folding in an environment like the *E. coli* periplasm, where probably most of the folding catalysts and chaperones, which normally act on the folding pathway in the ER lumen of the antibody producing B-cell, are absent. It is tempting to speculate that a consensus sequence defines a point in sequence space from which the observed sequences have diverged through genetic drift until the function of the protein is no longer maintained. This speculation is supported by experiments (Steipe *et al.*, 1994), where a clear correlation between degree of deviation from the consensus sequence and loss of thermodynamic stability of a murine V_L domain was found. Recent studies (Wörn & Plückthun, 1999) have shown

that, taking all available information into account, very stable and well-expressing antibodies can be engineered, suggesting that the β-sandwich framework is, in principle, a highly stable scaffold. Yet, most antibodies have diverged far from this point, first in the course of gene duplication during evolution of the locus, then in the V(D)J rearrangement where unfavorable CDR3s may be introduced, and finally in the somatic mutations, yielding antibody domains of very marginal biophysical integrity. The mouse repertoire is thought to be significantly larger than the human repertoire (Almagro *et al.*, 1998) and thus more deviations from the optimum are genetically encoded, partially explaining the difficulties in expressing antibody fragments derived from murine hybridomas. Several residues experimentally shown to be non-optimal (Spada *et al.*, 1998; Proba *et al.*, 1998) have been found to be encoded in some of the mouse germline genes, but in none of the human genes. Such residues are totally avoided in the present design.

Affinity maturation

The synthetic HuCAL genes were designed to contain unique restriction sites flanking the regions encoding the antigen binding loops, thereby making all six CDR regions accessible for diversification. The resulting modular gene structure, in combination with pre-built CDR library cassettes, will allow the rapid randomization of each CDR loop. We have constructed trinucleotide-based LCDR1 and HCDR2 cassettes using a design procedure identical with that described here for CDR3 cassettes (unpublished results). Hence an iterative randomization procedure can be envisaged, where the pool of binding sequences obtained after initial library selections can serve as starting material for the next iteration. Such a protocol would mimic the process of affinity maturation by somatic hypermutation observed during the natural immune response, even though the mechanism for achieving this would be different. It may be reasoned that this will be more efficient, as more of the mutations will be targeted to the region of interest. So far, the CDR walking process has been time-consuming, since the protocols and the CDR libraries had to be established for each individual antibody sequence. By using cassettes and the conserved restriction sites of the synthetic genes, however, an optimization of pools is possible, and the procedure is much more convenient. It has been shown now by several groups that the process of CDR walking, i.e. the iterative randomization of CDRs followed by stringent selection protocols, improved binding affinity of distinct antibody sequences dramatically (Yang *et al.*, 1995; Schier *et al.*, 1996b; Barbas & Burton, 1996; Rosok *et al.*, 1998; Wu *et al.*, 1998), and intrinsic affinities in the picomolar range could be obtained by this approach.

Nevertheless, framework residues can have indirect effects on binding by affecting the CDR conformations (Foote & Winter, 1992; Saul & Poljak, 1993), and a complete refinement may have to include these regions as well, e.g. by gene shuffling (Patten *et al.*, 1997) or ribosome display (Hanes *et al.*, 1998). Recently, the latter approach has been applied to the HuCAL1 library, and binders with sub-nanomolar affinities to several antigens have been obtained that do carry further mutations introduced by PCR (unpublished results).

Trinucleotide mixtures for CDR libraries

Using the 49 combined HuCAL frameworks, the initial libraries were created by randomizing two of the six CDR regions using trinucleotide building blocks. Sondek & Shortle (1992) first reported the use of a mixture of two trinucleotide phosphoramidites, but found a coupling yield of only 4% and large differences of relative coupling ratios. Virnekäs *et al.* (1994) showed that coupling of trinucleotide mixtures can be achieved with coupling yields as high as 96-98.5%, by carefully excluding traces of water during preparation of the phosphoramidite mixtures for coupling. However, in a first experiment using eight different trinucleotides, the individual codons were introduced with different frequencies (between one and 15 times within 63 positions being sequenced). No further improvement has been reported by other groups using similar building blocks (Lyttle *et al.*, 1995; Ono *et al.*, 1995; Kayushin *et al.*, 1996). Braunagel & Little (1997) used the trinucleotides described by Kayushin *et al.* (1996) in their approach to create a single-framework antibody library. However, no sequencing results were given to show the quality of the starting library or the distribution of individual codons.

We found that mixtures of trinucleotide phosphoramidites can be coupled in excellent yields. Oligonucleotides with a length of more than 100 bases and containing ten to 15 randomized positions have been successfully synthesized. Furthermore, no bias was found in most cases and trinucleotide-directed mutagenesis appears now to be the method of choice to achieve full control over the variability.

The option of using sub-stoichiometric coupling steps during oligonucleotide synthesis opens up a novel way of creating diversity by sequence and by length variation in a single oligonucleotide. We used sub-stoichiometric coupling for the generation of Vλ and V_H CDR3 libraries, and indeed it was possible to create CDR3s of different length with this method. However, the distribution of different length variants was in all cases shifted to shorter library members than calculated, suggesting that the stepwise coupling yields calculated from measuring the concentration of trityl cations, cleaved off the 5′-end, is higher than the actual coupling yield, i.e. the percentage of oligonucleotide chains being elongated during the sub-stoichio-

metric step. However, the parameters influencing the outcome of the sub-stoichiometric approach have not been studied in detail.

We decided initially to start with a diversification of CDR-L3 and CDR-H3, to imitate natural antibody generation. During the natural process of initial antibody generation, which results from genome rearrangements in the developing B-cell, most of the initial diversification is located in the V_H CDR3 region (VDJ-joining) and, to a lesser extent, the V_L CDR3 (VJ-joining). In the 3D structures of antibodies, both CDR3s form the so-called inner ring of the antigen binding site, and most of the antigen contacts are formed by residues located there (see Figures 1 and 3).

Comparison to semi-synthetic antibody libraries

The use of defined frameworks as the basis for generating an antibody library has been described before. Initial work on randomizing just CDR-H3 (Barbas *et al.*, 1992) has since then been extended to Vκ CDR3 (Barbas *et al.*, 1993; Yang *et al.*, 1995; Söderlind *et al.*,1995) or to single frameworks with all CDRs being randomized (Hayashi *et al.*, 1994; Iba & Kurosawa, 1997). Furthermore, sets of V_H genes extended with PCR primers that encode CDR-H3 libraries have been combined with a single V_L gene (Nissim *et al.*, 1994), or a limited set of V_L genes (De Kruif *et al.*, 1995), or a randomized repertoire of V_L genes (Griffiths *et al.*, 1994).

Most of the semi-synthetic human antibody libraries constructed so far focussed on exclusively randomizing CDR3s. For V_H, in most approaches A_{H93}, R_{H94} and D_{H101} were kept constant, and positions H95 to H100*z*, and usually H102 as well, were randomized (Hoogenboom & Winter, 1992; Barbas *et al.*, 1993, 1994; De Kruif *et al.*, 1995). The length of the CDR3s varied between six and 20 residues, with a preference for loops with six to 14 amino acid residues. De Kruif *et al.* (1995) constructed a set of eight CDRs between eight and 17 residues long, comprising completely and semi-randomized stretches.

For Vκ CDR3, usually residues L92 to L96 were randomized (Barbas *et al.*, 1993, 1994; Yang *et al.*, 1995; Söderlind *et al.*, 1995). The length of the CDR3s varied between seven and ten residues. Similarly, Hayashi *et al.* (1994) randomized 11 residues (including residue L97 of framework 4) of Vλ CDR3 in their approach to construct a one-framework library with all six CDRs being randomized. In contrast, Griffiths *et al.* (1994) used a whole set of 21 Vλ (as well as Vκ) germline genes and added, *via* PCR, specific CDR sequences comprising zero to five randomized codons. In all cases, codons were randomized by using mixtures of mononucleotides during oligonucleotide synthesis.

In our CDR3 design, we had to decide whether to stay close to the encoded variety with a preference for sequences actually found in selected antibodies or whether to follow a more daring approach. While, technically, both approaches are equally feasible, as it would depend only on the types of cassettes used, we opted to first examine CDR3 libraries close to the encoded variety. Even in a loop of this size, many combinations will be non-functional, and we wanted to secure a very high number of initial functional molecules. As library selection technology progresses, e.g. by the use of methods such as ribosome display (Hanes & Plückthun, 1997; Hanes *et al.*, 1998), much larger libraries will be screenable, and a larger set of variants may be simultaneously present, including those with structural defects.

When using the known rearranged sequences as a guide, it becomes an important question to what degree they represent "frozen accidents", explainable only by their evolutionary ancestry both at the germline and somatic level, or whether they are truly positively selected or are even due to genetic hotspots, encoded into the DNA sequence. The processes underlying somatic hypermutation are still not well understood. It was shown that heterologous genes replacing V gene segments undergo hypermutation *in vivo* as well (Yélamos *et al.*, 1995), and therefore it seems very unlikely that the V genes themselves determine at the genetic level where hypermutation occurs. A more reasonable explanation would be that selection determines which mutations finally survive. Various efforts have addressed this question (see, for example, Dorner *et al.*, 1998).

Weighing all arguments, we decided to take the natural distribution as our starting point. The modular approach permits any desired optimization strategy to be readily be carried out, once primary binders have been obtained, such as the introduction of V_L CDR1 and/or V_H CDR2 cassettes into single binders, or even pools of binders, since the sequences share identical restriction endonuclease sites adjacent to the CDRs. It would be also easily possible, for example, to keep the CDR3s of the selected pool of primary binders constant and shuffle V_H frameworks with randomized CDR2s. Alternatively, new sets of CDR3 libraries can be designed based on sequence motifs identified in the pool of primary binders. Furthermore, chain shuffling or even shuffling of elements such as CDRs or frameworks can now performed by restriction digest and religation.

Since HuCAL is fully synthetic, it is always possible to control the individual steps by analyzing the restriction pattern of individual clones or by sequencing, with artifacts being easily identified, whereas an immune repertoire cloned *via* PCR is more or less a black box.

By these means, searching the sequence space of human antibodies will be much faster and more efficient than by using the conventional approaches. Finally, we expect that the careful analysis of selected sequences will contain a wealth of structural information that can flow into subsequent versions of the library.

Conclusions and perspective

The HuCAL concept is based on covering the essential features of the human antibody repertoire with a minimal number of different sequences, which are designed to facilitate extensive manipulation with standard protein engineering techniques. The 49 combinations of master genes have been cloned as scFv genes in both orientations and as F_{ab} genes. Other formats like Fv fragments stabilized for example by disulfide-bridges (Glockshuber *et al.*, 1990; Brinkmann *et al.*, 1995; Rodrigues *et al.*, 1995) or fragments without any disulfide bonds (Wörn & Plückthun, 1998) useful for intrabody approaches (Cattaneo & Biocca, 1999; Wörn *et al.*, 2000) are easily adaptable and can be analyzed on the level of the master genes before actual library generation. Libraries can be rapidly created by inserting pre-built CDR cassettes into each of the 49 genes either separately or as mixed sequence pool, and the analysis of binding variants is facilitated by the fact that only small regions in the sequence are varied and that the three-dimensional models of all master frameworks have been built. It may therefore be possible for the first time to investigate experimentally why nature has evolved the distinct structural motifs found in the human antibody repertoire, and whether there are correlations of antibody structure with antigen class, antibody affinity and specificity. Future versions of HuCAL may therefore be enriched with antigen-type specific features.

Materials and Methods

Bacterial strains, phages, vectors

Molecular cloning was carried out using the *E. coli* strains JM83 (Yanisch-Perron *et al.*, 1985), XL1-Blue (Stratagene) or Top10 (Invitrogen). For expression experiments, JM83 was used. Phage-display libraries were generated and propagated using *E. coli* TG1 as host strain and M13K07 or VCSM13 as helper phage (all from Stratagene). The products from gene synthesis were cloned in pZero-1 (Invitrogen) or pCR-Script SK(+) (Stratagene) for sequencing. The pBS vector series used for antibody cloning and for expression analysis is a derivative of the phage-display vector pAK100 (Krebber *et al.*, 1997). The vector pBS10 contains the mature *bla* gene preceded by a region encoding the *ompA* signal sequence, a FLAG tag and an *Eco*RI cloning site between the *Xba*I/*Hin*dIII cloning sites of pAK100. The pBS10 vector was modified as follows in order to allow assembly of synthetic antibody genes. First, an oligonucleotide cassette encoding a synthetic *phoA* signal sequence (created by annealing the oligonucleotides O5phoA and O3phoA; all oligonucleotides constructed during this work are given in Table 2 of the Supplementary Material) was inserted into the *Xba*I/*Eco*RI sites. The resulting construct was designated pBS11. This *phoA* gene fragment contained a unique *Sap*I site, which was later used for insertion of V_H genes for the generation of F_{ab} fragments. The *phoA* gene fragment was extended by inserting a cassette created by annealing the oligonucleotides O5phoA_F and O3phoA_F into pBS11 *via* *Sap*I/*Eco*RI, thereby introducing the short improved FLAG tag (IDYKDE; Knappik & Plückthun, 1994). The resulting vector, designated pBS12, was later used for the assembly of scFv genes in the H-L orientation as well as for expression analysis. Second, the *Xba*I/*Eco*RI fragment from pBS10 was replaced by a cassette created by annealing the oligonucleotides O5stII and O3stII, thereby introducing a *stII* signal sequence containing a unique *Nsi*I site, which was later used for cloning of V_L genes for the generation of F_{ab} fragments. The resulting vector was designated pBS13. The *stII* gene fragment was extended by inserting a cassette created by annealing the oligonucleotides O5stII_F and O3stII_F into pBS13 *via* *Nsi*I/*Eco*RI, thereby introducing the short improved FLAG tag. The resulting vector, designated pBS14, was later used for the assembly of scFv genes in the L-H orientation. Vector pBS13b was constructed by removing the *Msc*I site in the *cat* resistance marker gene. The phage display vector pIG10.3 is a derivative of pIG10 (Ge *et al.*, 1995), where the first 249 codons of the mature full-length gene III were deleted. Briefly, the *Eco*RI/*Hin*dIII restriction fragment in the phagemid pIG10 was replaced by the c-myc tag for detection with the monoclonal antibody 9E10 (Munro & Pelham, 1986) followed by an amber codon and the truncated version of the gene III through PCR mutagenesis. The construction of the pMorph vector series, which is compatible with the HuCAL restriction sites and which was used for library cloning, will be described elsewhere (unpublished results). All vectors were constructed using site-directed mutagenesis (Kunkel, 1985), recursive PCR (Prodromou & Pearl, 1992) and overlap-extension PCR (Ge & Rudolph, 1997), and all constructs were subsequently verified by DNA sequencing (SequiServe, Vaterstetten, Germany).

Collection of human antibody sequences

Functional human germline sequences were downloaded from Genbank (Benson *et al.*, 1997), from the Kabat database† and from Vbase‡. Rearranged sequences were downloaded from Genbank and from the Kabat database. Kabat dump files were downloaded, variable domain amino acid sequences extracted and converted to the one-letter code. Sequences less than 90% complete or containing multiple undetermined residues in the regions of interest were eliminated. The automatic alignment generated by the program Pileup (Wisconsin Package, Version 8.1, 1995, Genetics Computer Group, Madison, WI, USA) was manually corrected to shift the gaps to the closest positions where they could be accommodated in the three-dimensional structure. The sequence files were converted and imported into Microsoft Excel®, where all subsequent alignments and analysis procedures took place. All alignments, numbering and loop regions (L1-L3, H1-H3) are according to structural criteria defined by Chothia and colleagues (see Chothia *et al.*, 1992; Tomlinson *et al.*, 1995; Williams *et al.*, 1996). CDRs were labeled as described by Kabat *et al.* (1991), even though this does not always correspond to the structural definition. Amino acid sequences are given in the single letter code according to standard IUPAC nomenclature. Germline sequences are named according to accepted locus nomenclature for each segment (Giudicelli *et al.*, 1997).

† ftp://ttwu.bme.nwu.edu/pub/database

‡ http://www. mrc-cpe.cam.ac.uk/imt-doc

Statistical analysis of the coverage of sequence space

After alignment and numbering according to Kabat, the databases were normalized by checking for multiple entries of closely related sequences, which we thought would indicate an artificial bias towards a specific set of rearranged sequences. Subsequently, the rearranged and the germline sequences were grouped into the various subfamilies. To assign the nearest germline to each rearranged sequence, the number of identities of a given rearranged sequence to each germline sequence was scored from position 1 to 92 (V_H) or position 1 to 95 (Vκ) or 1 to 95B (Vλ). If the result was ambiguous, e.g. the rearranged sequence was equidistant from two or more germline sequences, or if the best hit gave less than 80% identity, indicating either a very high level of somatic mutations or the origin from an at the time unknown germline gene, the rearranged sequence was omitted from the analysis.

By this analysis, the subfamilies that are used frequently by the human immune system were identified. The databases of rearranged sequences were used to calculate a consensus sequence for each frequently used subfamily. This was done by counting the number of amino acid residues used at each position (position variability) and subsequently identifying the amino acid residue most frequently used at each position. The consensus sequences were cross-checked with the consensus of the germline families to see whether the rearranged sequences were biased at certain positions towards amino acid residues that do not occur in the collected germline sequences, but this was found not to be the case. Subsequently, the CDR1 and CDR2 regions of the consensus sequences were replaced with the corresponding regions of the germline sequences that were most frequently used by the human immune system. For the framework 4 region, the consensus of all rearranged sequences was chosen. For each of these consensus sequences, the most homologous rearranged sequences were then identified and used for validating the consensus by identifying all framework residues that differed between the consensus and the most homologous rearranged sequences. These residues were regarded as artificial and checked by two means: first, the local context of the artificial residue was compared with the corresponding stretch of all the rearranged sequences in the database; and second, the long-range interactions of amino acid residues at these positions were analysed. To this end, the structures of human antibodies available from the Brookhaven Protein Database were analyzed, and the contacts of all side-chains were tabulated. If a certain artificial residue in the consensus sequence was found in the local context of rearranged sequences, and if this residue was not involved in side-chain interactions according to the structural analysis, it was kept at this position. Otherwise, the next most common residue was chosen and analyzed as described above. Finally, the consensus sequences were compared to the corresponding germline sequences and the number of differences were tabulated.

† www.biochem.ucl.ac.uk/ ~ roman/procheck/procheck.html

‡ ftp://ftp.ebi.ac.uk/pub/databases/codonusage/eco.cod

§ http://www.neb.com/ftp_info/rebase/

Molecular modeling

The structures of the HuCAL domains were predicted by homology modeling using the Homology, Biopolymer and Discover modules of the program InsightII version 95 (Biosym/MSI, San Diego, CA). To align different templates for the comparison of their conformation, a least-squares fit of the C^{α}-positions of residues H3-H7, H19-H23, H34-H40 (gapped according to structural criteria, not according to Kabat), H44-H50, H67-H71, H78-H82, I88-H94 and H102-H108 (V_H) or L3-L7, L20-L24, L33-L39, L43-L49, L62-L66, L71-L75, L84-L90 and L97-L103 (Vλ) was performed. The experimental structures displaying the highest degree of sequence similarity to the different HuCAL constructs are listed in Table 1 of the Supplementary Material. Structural differences between these templates were analyzed to identify the sequence differences responsible for the deviations. The conformation of the dummy CDR3 s was taken from the structure of the humanized 4D5 version 8 (PDB entry 1FVC). Coordinates were assigned using the Homology module and the resulting models checked for steric clashes and cavities before energy minimization (module Discover, CFF91 forcefields). The stereochemical quality of the final domain models was evaluated with the program PROCHECK† (Laskowski *et al.*, 1993; Morris *et al.*, 1992).

Gene synthesis and assembly

Consensus amino acid sequences were back-translated into DNA sequences using the GCG software package (Genetics Computer Group, Madison, WI, USA) and a Codon definition file that included only the codons that are used frequently in *E. coli*‡. All possible silent (and commercially available) restriction sites based on version 501 of the REBASE list of restriction enzymes (Roberts & Macelis, 1999§) were subsequently identified in the resulting DNA sequences and tabulated. These tables were used to identify all cleavage sites that were located close to the CDR and framework borders, and that could be introduced into all genes of the three classes (V_H, Vκ or Vλ) simultaneously at the same position. Further editing was done as described in Results. For each of the 14 resulting genes, six overlapping oligonucleotides were designed. Since both the CDR3 and the framework 4 gene segments were identical in all Vκ, Vλ and V_H genes, respectively, this part was constructed only once in each case. The region of overlap was chosen to give a theoretical t_m of 58°C (corresponding to a ΔG of about −20 kcal/mol), and the 3' nucleotide was chosen to be either C or G. The design was examined and optimized in terms of potential stem-loop formation, dimer formation and potential unspecific hybridization sites with all other oligonucleotides (duplex formation) using the VectorNTI® software (Informax, Inc.). PCR assembly (Prodromou & Pearl, 1992) was performed by mixing 200 pmol of each of the oligonucleotides in a 100 µl reaction volume containing 20 nmol of dNTPs and five units of Pfu polymerase (Stratagene). After a first cycle with three minutes at 94°C, two minutes at 60°C and one minute at 72°C using a hotstart procedure, 31 PCR cycles were performed (one minute at 94°C, two minutes at 60°C and one minute at 72°C), the products were purified using the QIAgen PCR purification kit and blunt-end ligated with either the pCR-Script KS(+) (cut with *Sfr*I) or the pZero-1 vector (cut with *Eco*RV).

Insert containing clones were screened by blue-white selection (pCR-Script KS(+)) or directly picked (pZero-1) and sequenced.

The seven synthesized V_H genes covered the sequences from the first unique 5′ restriction site located in the *phoA* signal sequence region (*SapI*) to the last unique 3′ restriction site located in the framework 3 region prior to the CDR3 (*BssHII*). All genes were synthesized with their authentic N termini and without the short FLAG sequence, which was added later during the construction of scFv display vectors. The heavy chain C_H1 domain (subtype IgG1, Genbank accession number A49444) including the V_H framework 4 region was assembled using eight oligonucleotides (OCH1 to OCH8) and inserted into pCR-Script KS(+). The C_H1 gene sequence was designed for optimal *E. coli* codon usage. Additionally, restriction sites for *SalI* and *EcoRI* were incorporated at the 5′- and 3′-ends, respectively, and most internal restriction sites were removed during the gene design. In a second step, the V_H dummy CDR3 region was inserted as a *PstI*/*StyI* cassette using the oligonucleotides OHCDR3P and OHCDR3 M. The V_H gene fragments (*SapI*/*BssHII*) were assembled with the CDR3-framework 4-C_H1 sequence (*BssHII*/*EcoRI*) by a three-fragment ligation with the vector pBS11 (*SapI*/*EcoRI*), yielding seven Fd fragments for construction of F_{ab} expression vectors. The N-terminal FLAG tag was added later for scFv constructions by cloning the Fd fragments into the vector pBS12 using the restriction enzymes *MfeI* and *EcoRI*.

The four synthesized V_k kappa genes covered the sequences from the unique 5′ restriction site located in the *stII* signal sequence region (*NsiI*) to the unique 3′ restriction site located in the framework 3 region prior to the CDR3 (*Eco57I*). The human kappa constant domain Cκ (Genbank accession number P01834) including the Vκ framework 4 region, the V_L dummy CDR3, and part of the Vκ framework 3 region (including the *Eco57I* restriction site) was synthesized using eight oligonucleotides (OCLk1 to OCLk8) and inserted into pCR-Script KS(+). The Cκ gene sequence was optimized for *E. coli* codon usage, the internal restriction sites except *AccI* were removed and the restriction sites for *BsiWI* and *StuI* were incorporated at the 5′ and 3′-end, respectively. The Vκ gene fragments (*NsiI*/*Eco57I*) were then assembled with the CDR3-framework 4-Cκ sequence (*SphI*/*Eco57I*) by a three-fragment ligation with the vector pBS13 (*SphI*/*NsiI*), yielding four kappa light chain fragments for construction of F_{ab} expression vectors.

The three synthesized V_L lambda genes covered the sequences from the unique 5′ restriction site located in the *stII* signal sequence region (*NsiI*) to the unique 3′ restriction site located in the framework 3 region prior to the CDR3 (*BlsI*). All genes were synthesized using their authentic N termini, i.e. without the aspartate-isoleucine stretch encoded by an *Eco*RV site used for the Vκ genes. The human lambda constant domain Cλ1 (Genbank accession number P01842) including the Vλ framework 4 region, the V_L dummy CDR3, and part of the Vλ framework 3 region was assembled as *BbsI*/*SphI* fragment by complete gene synthesis with 12 oligonucleotides. The Vλ gene fragments (*NsiI*/*BbsI*) were assembled with the CDR3-framework 4-Cλ sequence (*BbsI*/*SphI*) by a three-fragment ligation with the vector pBS13b (*SphI*/*NsiI*), yielding three lambda light chain fragments for construction of F_{ab} expression vectors. In order to assemble V_L-V_H scFv vectors, the Vλ gene fragments were further modified: the Vλ gene fragments were PCR amplified from pBS13b using the forward primers OLxFw1DIP (where × denotes the Vλ sub-family) and the backward primer OLFw4 M, and the PCR products were blunt-end ligated into pBS14_scx1H3 (a V_L-V_H scFv expression vector constructed as described below), which had been cut with *EcoRV*/*BsiWI* and made blunt-ended by teatment with S_1 nuclease. The resulting three plasmids were named pBS14_scλ$_{(1\text{-}3)}$H3 and contained Vλ genes where the two N-terminal codons had been changed to the *Eco*RV recognition sequence encoding aspartate-isoleucine, in order to allow the same scFv cloning protocol as used for the Vκ genes. These plasmids were used for assembly of the V_L-V_H scFv expression vectors (see below). In order to assemble V_H-V_L scFv vectors, a cassette constructed by annealing the oligonucleotides OLEco5 and OLEco3 was inserted into the lambda light chain containing vectors pBS13b_Vλ$_{(1\text{-}3)}$Cλ cut with *MscI*/*EcoRI*, thereby replacing the C_L constant domain gene fragment by an in-frame *EcoRI* site. The resulting three plasmids were designated pBS13b_Vλ$_{(1\text{-}3)}$_E. After cutting these vectors with *XmaI*/*HindIII*, the 3′-region of each the three Vλ genes including the in-frame *EcoRI* sequence were isolated and inserted into the corresponding pBS14_scλ$_{(1\text{-}3)}$H3 vectors, thereby adding the 5′ *EcoRV* and the 3′ *EcoRI* sites to each of the three Vλ genes. These genes were used to assemble the V_H-V_L scFv expression vectors (see below).

F_{ab} expression plasmids were constructed by combining each of the heavy chain Fd fragments cut with *SphI*/*EcoRI* and each of the light chain fragments cut with *XbaI*/*SphI* with the pBS13 vector cut with *XbaI*/*EcoRI* in a three-fragment ligation reaction. The 49 resulting plasmids were verified by restriction enzyme digestions. Here, the Vλ gene fragments contain their authentic N termini, and there is no FLAG tag sequence attached to the antibody F_{ab} genes.

The scFv expression plasmids in the orientation V_L-V_H were constructed as follows: the Cκ gene fragment from pB13_Vk2Cκ was removed by cutting the plasmid with *BsiWI*/*SphI* and replaced by an oligonucleotide cassette encoding a 20 amino acid residue linker plus the additional restriction sites *MfeI* and *EcoRI* for later insertion of the V_H genes. The cassette was constructed by annealing the oligonucleotides OLHLiP and OLHLiM. Subsequently, the remaining Vκ and Vλ genes were inserted as *XbaI*/*BsiWI* fragments and the V_H genes were inserted as *MfeI*/*EcoRI* fragments.

The 49 scFv expression plasmids in the orientation V_H-V_L were constructed as follows: the CH1 gene fragment from pBS12_VH3CH1 was removed by cutting the plasmid with *BlpI*/*EcoRI* and replaced by an oligonucleotide cassette encoding a 20 amino acid residue linker plus the additional restriction site *EcoRV* for later insertion of the V_L genes. The cassette was constructed by annealing the oligonucleotides OHLLiP and OHLLiM. Subsequently the Vκ and Vλ genes were inserted as *EcoRV*/*EcoRI* fragments and the V_H genes were inserted as *XbaI*/*BlpI* fragments. These 49 vectors were used for expression analysis, and the scFv genes were later used for library construction.

Expression analysis

Growth curves and expression data were obtained essentially as described (Knappik & Plückthun, 1995). Briefly, *E. coli* JM83 cultures containing the appropriate

scFv expression vectors were grown at 30 °C and induced with 1 mM IPTG. After two hours of expression, cells were harvested, normalized to identical absorbance, lysed and separated into soluble and insoluble cell fractions by centrifugation. The fractions were assayed by reducing SDS-PAGE, blotting and immunostaining using the anti-FLAG antibody M1 (Sigma), and the amount of soluble and insoluble scFv protein produced was quantified densitometrically. The scFv gene H3x2 was used as internal control in each expression experiment.

Expression kinetics were measured as follows: *E. coli* JM83 cells were transformed with the scFv genes cloned in the expression vector pMorphx7_FS (unpublished results) and grown in 1 l shaking-flask cultures at 30 °C. After induction with 1 mM IPTG, 50 ml of culture was harvested each hour, the cells were normalized to $A = 50$, lysed by sonification, and the crude extracts were stored at −20 °C. After ten hours induction, the remaining culture (500 ml) was harvested, lysed, the scFv fragment was purified using Poros Streptactin affinity chromatography (IBA, Göttingen, Germany), and the amount of purified scFv was determined. The functional scFv expression yield at the different time-points was then determined by ELISA measurements, where the purified antibody fragment of known concentration served as internal standard used to calculate the scFv amount based on the ELISA signal obtained.

CDR analysis and library design

The aligned collections of rearranged human antibody V_H and V_L sequences were used for analysis of CDR3 length and composition. For analysis of V_H CDR3 s, all sequences were grouped together because sequence alignments are not possible in this highly diverse region. For Vκ and Vλ CDR3s, the subfamilies were analyzed separately. Within the individual alignments, the CDRs were grouped according to CDR length. Assignment of the individual groups to canonical structures was done according to the rules described by Chothia *et al.* (1989). All analysis was done using Microsoft Excel®.

Synthesis of trinucleotide-containing oligonucleotides

Synthesis of *O*-methyl trinucleotide phosphoramidites and their application in automated DNA synthesis has been described (Virnekäs *et al.*, 1994). Trinucleotide mixtures were prepared by mixing appropriate stoichiometric amounts of solid phosphoramidites, assuming equal reactivities of all 20 trinucleotides. The mixtures were dried under argon and dissolved to yield 0.1 M solutions as described (Virnekäs *et al.*, 1994). Automated synthesis was performed on an Applied Biosystems DNA synthesizer 380B. The synthesis reagents were obtained from Applied Biosystems and MWG (Ebersberg, Germany). All trinucleotide-based syntheses were performed on columns with polystyrene support, 1000 Å, 40 nmol (Applied Biosystems, art. 401072 to 401075). For synthesizing stretches with mononucleotide building blocks of the oligonucleotides, conventional mononucleotide *O*-cyanoethyl phosphoramidites, and the standard synthesis cycle SSCEAF (single coupling, 15 seconds wait step) were used. When coupling trinucleotide mixtures stoichiometrically, the standard cycle was changed to double couple, including a 100 seconds wait step after the first, and a 400 seconds wait step after the second coupling. For sub-stoichiometric couplings, the time for delivering activated phosphoramidite solution to the column was reduced to achieve approximately 50 % coupling yield. If substoichiometric coupling rates were much higher or lower than 50 %, either the time was adjusted for the subsequent couplings to obtain an average yield of 50 % over all substoichiometric couplings or an additional substoichiometric coupling step was added. Deprotection of the oligonucleotides was performed as described (Virnekäs *et al.*, 1994). All trinucleotide-containing oligoncleotides synthesized for CDR3 library generation are given in Table 3 of the Supplementary Material.

Cassette preparation

The oligonucleotides were resuspended in TE buffer and purified with an S200 column (Pharmacia) according to the supplier's manual. The complementary strand was synthesized with Klenow polymerase (New England Biolabs). Approximately 5 nmol of oligonucleotide was mixed with a cassette-specific corresponding primer at a ratio of 1:1.2, respectively, heated for ten minutes to 80 °C followed by slowly cooling to room temperature: 10 µl of a 10 mM dNTP mixture, 15 µl of Klenow buffer, 2 µl of Klenow polymerase and water to 150 µl final volume were added. The fill-in reaction was performed at 37 °C for two hours and purified with a Nick Spin column according to the supplier's manual (Pharmacia Biotech). The fill-in reaction was checked by an analytical FMC agarose gel (Biomol). To amplify the fill-in products, PCR reactions were performed using 1 µl of the fill-in reaction mixtures (approximately 25 pmol) and 100 pmol of each primer (fill-in primer plus second cassette-specific primer) in each case (30 cycles, one minute at 94 °C; one minute at 54 °C, one minute at 72 °C). The PCR mixtures were purified with a Nick Spin column. The oligonucleotide library cassettes were prepared for ligation by adding 30 µl of buffer to 100 µl of the purified PCR product, 150 units of each of the corresponding restriction enzymes, and water to a final volume of 300 µl, and by digesting overnight at 37 °C. The cassettes were purified on 4 % FMC agarose gels (Biomol), and recovered from the gel *via* BIOTRAP elution (Schleicher & Schuell, Germany) according to the supplier's manual (approximately two hours at 100 V/50-70 mA). The solutions containing the cassettes were desalted with Nick spin columns. The quality of the cassettes was checked by analytical FMC agarose gels (4 %).

Generation of the HuCAL1 library

Template V_H vectors were created by inserting the seven HuCAL V_H master genes as Fd genes from the vector pBS13 into the display vector pMorph7 (unpublished results). The V_H CDR3 sequences were then replaced by a 1220 bp dummy fragment containing the β-lactamase gene, thereby facilitating subsequent steps for vector fragment preparation. The template V_H vectors were cut with *StyI*/*Hin*dIII to remove the CH1 gene fragment, and the vector fragments were purified. At this step, the two template vector fragments encoding the V_H1A and V_H1B master genes were mixed in an equimolar ratio, resulting in six V_H vector templates.

Template V_L vectors were constructed by first inserting the HuCAL scFv master genes containing the

HuCAL H3 gene in combination with all seven V_L genes into the vector pMorph4, and second, replacing the V_L CDR3 sequences by the β-lactamase gene dummy sequence as described above. The resulting seven template V_L vectors were then purified, the 1220 bp dummy fragment was removed by cutting with *BbsI/MscI* for Vκ gene-containing vectors and *BbsI/HpaI* for the Vλ gene-containing vectors. The prepared trinucleotide cassettes encoding the V_L CDR3 libraries were then ligated separately with the seven V_L template vector fragments (25 fmol of each vector was ligated with 250 fmol of each cassette for three hours at room temperature), and the ligation mixtures were electroporated in 0.9 ml of *E. coli* TG1 cells, yielding altogether 1.14×10^7 independent colonies. The colonies were scraped off the selection plates, and the V_L CDR3 libraries were stored in 20% (w/v) glycerol at −80°C. Phagemid DNA of the four Vκ and the three Vλ libraries was prepared and two pools were created by mixing the four Vκ (κ_{mix}) and the three Vλ (λ_{mix}) libraries in an equimolar ratio. The two DNA pools were treated with *StyI/HindIII*, the V_L gene libraries were purified using agarose gel electrophoresis and 75 fmol of each pool was ligated with 25 fmol of each of the six V_H template vectors (see above), and electroporated in 0.3 ml of *E. coli* TG1 cells, resulting in altogether 2.3×10^7 colonies for the 12 library pools. In the final step, these 12 libraries were prepared as DNA, cut with *BssHII/StyI* to remove the β-lactamase dummy gene inside the V_H CDR3 region, and the two V_H CDR3 trinucleotide library cassettes (HCDR3a and HCDR3b) were inserted separately by ligation using the same conditions as above. After electroporation into 7.2 ml of *E. coli* TG1 (12 electroporations for each library), we obtained altogether 2.1×10^9 independent colonies. The diversity for each pool was between 0.6×10^9 ($V_H2^{\kappa}{}_{mix}$) and 3.9×10^8 ($V_H6^{\kappa mix}$). The colonies were scraped off the selection plates, and the 24 HuCAL1 library were stored as aliquots in 20% glycerol at −80°C.

Data Bank accession numbers

The coordinates of the 14 framework models have been deposited in the RCSB Protein Data Bank, entries 1DGX (Vκ1), 1DH4 (Vκ2), 1DH5 (Vκ3), 1DH6 (Vκ4), 1DH7 (Vλ1), 1DH8 (Vλ2), 1DH9 (Vλ3), 1DHA (VH1A), 1DHO (VH1B), 1DHQ (VH2), 1DHU (VH3), 1DHV (VH4), 1DHW (VH5) and 1DHZ (VH6).

Acknowledgments

We thank Ilona Meyer, Peter Rudolph, Martina Mayer, Isabel Rapp, Ursi Holzinger and Siegfried Hirler for excellent technical assistance, Leodevico Ilag for advice in structural analysis, and Titus Kretzschmar and William Reisdorf for critical reading of the manuscript. We thank Josef Hanes, Barbara Krebs, Titus Kretzschmar, Ralf Ostendorp, Josef Prassler, Christine Rothe, Silke Reiffert, Christiane Schaffitzel and Robert Schier for providing the data given in Table 4. We thank David Fischer, Pharmacia & Upjohn, for supplying us with purified Mac-1 and ICAM-1.

References

Al-Lazikani, B., Lesk, A. M. & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927-948.

Almagro, J. C., Hernandez, I., Ramirez, M. C. & Vargas-Madrazo, E. (1998). Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. *Immunogenetics*, **47**, 355-363.

Baca, M., Presta, L. G., O'Connor, S. J. & Wells, J. A. (1997). Antibody humanization using monovalent phage display. *J. Biol. Chem.* **272**, 10678-10684.

Barbas, C. F. & Burton, D. R. (1996). Selection and evolution of high-affinity human anti-viral antibodies. *Trends Biotech.* **14**, 230-234.

Barbas, C. F., Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992). Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl Acad. Sci. USA*, **89**, 4457-4461.

Barbas, C. F., Amberg, W., Simoncsits, A., Jones, T. M. & Lerner, R. A. (1993). Selection of human anti-hapten antibodies from semi-synthetic libraries. *Gene*, **137**, 57-62.

Barbas, S. M., Ghazal, P., Barbas, C. F., III & Burton, D. R. (1994). Recognition of DNA by synthetic antibodies. *J. Am. Chem. Soc.* **116**, 2161-2162.

Barbie, V. & Lefranc, M. P. (1998). The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp. Clin. Immunogenet.* **15**, 171-183.

Barre, S., Greenberg, A. S., Flajnik, M. F. & Chothia, C. (1994). Structural conservation of hypervariable regions in immunoglobulins evolution. *Nature Struct. Biol.* **1**, 915-920.

Baskin, B., Islam, K. B. & Smith, C. I. (1998). Characterization of the CDR3 region of rearranged alpha heavy chain genes in human fetal liver. *Clin. Exp. Immunol.* **112**, 44-47.

Benson, D. A., Boguski, M. S., Lipman, D. J. & Ostell, J. (1997). Genbank. *Nucl. Acids Res.* **25**, 1-6.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Boel, E., Bootsma, E., de Kruif, J., Jansze, M., Klingman, K. L., Vandijk, H. & Logtenberg, T. (1998). Phage antibodies obtained by competitive selection on complement-resistant *Moraxella* (*Branhamella*) *catarrhalis* recognize the high-molecular-weight outer membrane protein. *Infect. Immun.* **66**, 83-88.

Bothmann, H. & Plückthun, A. (1998). Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nature Biotech.* **16**, 376-380.

Braunagel, M. & Little, M. (1997). Construction of a semisynthetic antibody library using trinucleotide oligos. *Nucl. Acids Res.* **25**, 4690-4691.

Brinkmann, U., Chowdhury, P. S., Roscoe, D. M. & Pastan, I. (1995). Phage display of disulfide-stabilized Fv fragments. *J. Immunol. Methods*, **182**, 41-50.

Carter, P. & Merchant, A. M. (1997). Engineering antibodies for imaging and therapy. *Curr. Opin. Biotechnol.* **8**, 449-454.

Carter, P., Kelley, R. F., Rodrigues, M. L., Snedecor, B., Covarrubias, M., Velligan, M. D., Wong, W. L., Rowland, A. M., Kotts, C. E. & Carver, M. E. (1992a). High level *Escherichia coli* expression and

production of a bivalent humanized antibody fragment. *Bio/Technology*, **10**, 163-167.

Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B., Henner, D., Wong, W. L., Rowland, A. M., Kotts, C., Carver, M. E. & Shepard, H. M. (1992b). Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl Acad. Sci. USA*, **89**, 4285-4289.

Cattaneo, A. & Biocca, S. (1999). The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115-121.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901-917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877-883.

Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). Structural repertoire of the human VH segments. *J. Mol. Biol.* **227**, 799-817.

Chothia, C., Gelfand, I. & Kister, A. (1998). Structural determinants in the sequences of immunoglobulin variable domain. *J. Mol. Biol.* **278**, 457-479.

Cook, G. P. & Tomlinson, I. M. (1995). The human immunoglobulin V-H repertoire. *Immunol. Today*, **16**, 237-242.

Cox, J. P., Tomlinson, I. M. & Winter, G. (1994). A directory of human germ-line V kappa segments reveals a strong bias in their usage. *Eur. J. Immunol.* **24**, 827-836.

Cwirla, S. E., Balasubramanian, P., Duffin, D. J., Wagstrom, C. R., Gates, C. M., Singer, S. C., Davis, A. M., Tansik, R. L., Mattheakis, L. C., Boytos, C. M., Schatz, P. J., Baccanari, D. P., Wrighton, N. C., Barrett, R. W. & Dower, W. J. (1997). Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine. *Science*, **276**, 1696-1699.

Dall'Acqua, W. & Carter, P. (1998). Antibody engineering. *Curr. Opin. Struct. Biol.* **8**, 443-450.

De Haard, H. J., Van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P., De Bruine, A. P., Arends, J. W. & Hoogenboom, H. R. (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.* **274**, 18218-18230.

De Kruif, J., Boel, E. & Logtenberg, T. (1995). Selection and application of human single chain Fv antibody fragments from a semi-synthetic phage antibody display library with designed CDR3 regions. *J. Mol. Biol.* **248**, 97-105.

Deng, S. J., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R. & Narang, S. A. (1994). Selection of antibody single-chain variable fragments with improved carbohydrate binding by phage display. *J. Biol. Chem.* **269**, 9533-9538.

Dorner, T., Foster, S. J., Farner, N. L. & Lipsky, P. E. (1998). Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. *Eur. J. Immunol.* **28**, 3384-3396.

Dorsam, H., Rohrbach, P., Kurschner, T., Kipriyanov, S., Renner, S., Braunagel, M., Welschof, M. & Little, M. (1997). Antibodies to steroids from a small human naive IgM library. *FEBS Letters*, **414**, 7-13.

Foote, J. & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**, 487-499.

Forsberg, C., Forsgren, M., Jaki, M., Norin, M., Sterky, C., Enhörning, A., Larsson, K., Ericsson, M. & Björk, P. (1997). Identification of framework residues in a secreted recombinant antibody fragment that control production level and localization in *Escherichia coli*. *J. Biol. Chem.* **272**, 12430-12436.

Ge, L. & Rudolph, P. (1997). Simultaneous introduction of multiple mutations using overlap extension PCR. *Biotechniques*, **22**, 28 ff.

Ge, L., Knappik, A., Pack, P., Freund, C. & Plückthun, A. (1995). Expressing antibodies in *Escherichia coli*. In *Antibody Engineering* (Borrebaeck, C. A. K., ed.), pp. 229-266, Oxford University Press, Oxford.

Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. & Curtiss, R. (1997). Display of heterologous proteins on the surface of microorganisms-from the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotech.* **15**, 29-34.

Giudicelli, V., Chaume, D., Bodmer, J., Muller, W., Busin, C., Marsh, S., Bontrop, R., Marc, L., Malik, A. & Lefranc, M. P. (1997). IMGT, the International ImMunoGeneTics Database. *Nucl. Acids Res.* **25**, 206-211.

Glockshuber, R., Malia, M., Pfitzinger, I. & Plückthun, A. (1990). A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry*, **29**, 1362-1367.

Green, N. S., Lin, M. M. & Scharff, M. D. (1998). Somatic hypermutation of antibody genes: a hot spot warms up. *Bioessays*, **20**, 227-234.

Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. & Harrison, J. L., *et al.* (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245-3260.

Hanes, J. & Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA*, **94**, 4937-4942.

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Plückthun, A. (1998). Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc. Natl Acad. Sci. USA*, **95**, 14130-14135.

Hayashi, N., Welschof, M., Zewe, M., Braunagel, M., Dübel, S., Breitling, F. & Little, M. (1994). Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques*, **17**, 310, 312, 314-315.

He, Y. Y., Stockley, P. G. & Gold, L. (1996). *In vitro* evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ. *J. Mol. Biol.* **255**, 55-66.

Hoogenboom, H. R. & Winter, G. (1992). By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged *in vitro*. *J. Mol. Biol.* **227**, 381-388.

Hoogenboom, H. R., De Bruine, A. P., Hufton, S. E., Hoet, R. M., Arends, J. W. & Roovers, R. C. (1998). Antibody phage display technology and its applications. *Immunotechniques*, **4**, 1-20.

Iba, Y. & Kurosawa, Y. (1997). Comparison of strategies for the construction of libraries of artificial antibodies. *Immunol. Cell Biol.* **75**, 217-221.

Ignatovich, O., Tomlinson, I. M., Jones, P. T. & Winter, G. (1997). The creation of diversity in the human

immunoglobulin V-lambda repertoire. *J. Mol. Biol.* **268**, 69-77.

Jackson, J. R., Sathe, G., Rosenberg, M. & Sweet, R. (1995). In vitro antibody maturation-improvement of a high affinity neutralizing antibody against IL-1 beta. *J. Immunol.* **154**, 3310-3319.

Jirholt, P., Ohlin, M., Borrebaeck, C. A. K. & Söderlind, E. (1998). Exploiting sequence space-shuffling in vivo formed complementarity determining regions into a master framework. *Gene*, **215**, 471-476.

Johnson, G., Kabat, E. & Wu, T. T. (1996). Kabat database of sequences of proteins of immunological interest. In. *WEIR'S Handbook of Experimental Immunology I. Immunochemistry and Molecular Immunology* (Herzenberg, L. A., Weir, W. M., Herzenberg, L. A. & Blackwell, C., eds), 5th edit. chapt. 6, pp. 6.1-6.21, Blackwell Science Inc. , Cambridge, MA.

Jung, S. & Plückthun, A. (1997). Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng.* **10**, 959-966.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th edit., NIH Publication no. 91-3242, US Department of Health and Human Services, Washington, DC.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J. L., Wang, J. & Shimizu, N. (1997). One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* **7**, 250-261.

Kayushin, A. L., Korosteleva, M. D., Miroshnikov, A. I., Kosch, W., Zubov, D. & Piel, N. (1996). A convenient approach to the synthesis of trinucleotide phosphoramidites-synthons for the generation of oligonucleotide/peptide libraries. *Nucl. Acids Res.* **24**, 3748-3755.

Kieke, M. C., Cho, B. K., Boder, E. T., Kranz, D. M. & Wittrup, K. D. (1997). Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng.* **10**, 1303-1310.

Knappik, A. & Plückthun, A. (1994). An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. *Biotechniques*, **17**, 754-761.

Knappik, A. & Plückthun, A. (1995). Engineered turns of a recombinant antibody improve its *in vivo* folding. *Protein Eng.* **8**, 81-89.

Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R. & Plückthun, A. (1997). Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods*, **201**, 35-55.

Kunkel, T. A. (1985). Rapid & efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488-492.

Langedijk, A. C., Honegger, A., Maat, J., Planta, R. J., van Schaik, R. C. & Plückthun, A. (1998). The nature of antibody heavy chain residue H6 strongly influences the stability of a VH domain lacking the disulfide bridge. *J. Mol. Biol.* **283**, 95-110.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Lyttle, M. H., Napolitano, E. W., Calio, B. L. & Kauvar, L. M. (1995). Mutagenesis using trinucleotide beta-cyanoethyl phosphoramidites. *Biotechniques*, **19**, 274-281.

Matsuda, F. & Honjo, T. (1996). Organization of the human immunoglobulin heavy-chain locus. *Advan. Immunol.* **62**, 1-29.

Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. (1998). Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.* **275**, 269-294.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Struct. Funct. Genet.* **12**, 345-364.

Munro, S. & Pelham, H. R. (1986). An Hsp70-like protein in the ER: identity with the 78 kD glucose-regulated protein and immunoglobulin heavy chain binding protein. *Cell*, **46**, 291-300.

Nieba, L., Honegger, A., Krebber, C. & Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10**, 435-444.

Nissim, A., Hoogenboom, H. R., Tomlinson, I. M., Flynn, G., Midgley, C., Lane, D. & Winter, G. (1994). Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J.* **13**, 692-698.

Oliva, B., Bates, P. A., Querol, E., Aviles, F. X. & Sternberg, M. E. (1998). Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction. *J. Mol. Biol.* **279**, 1193-1210.

Ono, A., Matsuda, A., Zhao, J. & Santi, D. V. (1995). The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis. *Nucl. Acids Res.* **23**, 4677-4682.

Pallares, N., Frippiat, J. P., Giudicelli, V. & Lefranc, M. P. (1998). The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp. Clin. Immunogenet.* **15**, 8-18.

Patten, P. A., Howard, R. J. & Stemmer, W. P. C. (1997). Applications of DNA shuffling to pharmaceuticals and vaccines. *Curr. Opin. Biotechnol.* 8, 724-733.

Perelson, A. S. (1989). Immune network theory. *Immunol. Rev.* **110**, 5-36.

Pini, A., Viti, F., Santucci, A., Carnemolla, B., Zardi, L., Neri, P. & Neri, D. (1998). Design and use of a phage display library-human antibodies with subnanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. *J. Biol. Chem.* **273**, 21769-21776.

Proba, K., Worn, A., Honegger, A. & Plückthun, A. (1998). Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275**, 245-253.

Prodromou, C. & Pearl, L. H. (1992). Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* **5**, 827-829.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature*, 381, 751-758.

Roberts, R. J. & Macelis, D. (1999). REBASE-restriction enzymes and methylases. *Nucl. Acids Res.* **27**, 312-313.

Rodi, D. J. & Makowski, L. (1999). Phage-display technology-finding a needle in a vast molecular haystack. *Curr. Opin. Biotechnol.* **10**, 87-93.

Rodrigues, M. L., Presta, L. G., Kotts, C. E., Wirth, C., Mordenti, J., Osaka, G., Wong, W. L. T., Nuijens, A., Blackburn, B. & Carter, P. (1995). Development of a humanized disulfide-stabilized anti-beta P185(HER2) Fv-beta-lactamase fusion protein for activation of a cephalosporin doxorubicin prodrug. *Cancer Res.* **55**, 63-70.

Rosok, M. J., Eghtedarzadehkondri, M., Young, K., Bajorath, J., Glaser, S. & Yelton, D. (1998). Analysis of BR96 binding sites for antigen and anti-idiotype by codon-based scanning mutagenesis. *J. Immunol.* **160**, 2353-2359.

Saul, F. A. & Poljak, R. J. (1993). Structural patterns at residue positions 9, 18, 67 and 82 in the VH framework regions of human and murine immunoglobulins. *J. Mol. Biol.* **230**, 15-20.

Schier, R., Bye, J., Apell, G., McCall, A., Adams, G. P., Malmqvist, M., Weiner, L. M. & Marks, J. D. (1996a). Isolation of high-affinity monomeric human anti-c-erbb-2 single chain Fv using affinity-driven selection. *J. Mol. Biol.* **255**, 28-43.

Schier, R., McCall, A., Adams, G. P., Marhall, K. W., Merritt, H., Yim, M., Crawford, R. S., Weiner, L. M., Marks, C. & Marks, J. D. (1996b). Isolation of picomolar affinity anti-c-erbb-2 single-chain Fv by molecular evolution of the complementarity determining regions in the center of the antibody binding site. *J. Mol. Biol.* **263**, 551-567.

Searle, S. J., Pedersen, J. T., Henry, A. H., Webster, D. M. & Rees, A. R. (1995). Expressing antibodies in *Escherichia coli*. In *Antibody Engineering* (Borrebaeck, C. A. K., ed.), pp. 3-51, Oxford University Press, Oxford.

Sheets, M. D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J. C. & Marks, J. D. (1998). Efficient construction of a large nonimmune phage antibody library-the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl Acad. Sci. USA*, **95**, 6157-6162.

Shirai, H., Kidera, A. & Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Letters*, **399**, 1-8.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315-1317.

Smith, G. P. & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.* **217**, 228-257.

Söderlind, E., Vergeles, M. & Borrebaeck, C. A. K. (1995). Domain libraries: synthetic diversity for de novo design of antibody V-regions. *Gene*, **160**, 269-272.

Sondek, J. & Shortle, D. (1992). A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc. Natl Acad. Sci. USA*, **89**, 3581-3585.

Spada, S., Krebber, C. & Plückthun, A. (1997). Selectively infective phages (SIP). *Biol. Chem.* **378**, 445-456.

Spada, S., Honegger, A. & Plückthun, A. (1998). Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology: the kink in the first strand of antibody kappa domains. *J. Mol. Biol.* **283**, 395-407.

Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188-192.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.

Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline VH sequences reveals about 50 groups of VH segments with different hypervariable loops. *J. Mol. Biol.* **227**, 776-798.

Tomlinson, I. M., Cox, J. P., Gherardi, E., Lesk, A. M. & Chothia, C. (1995). The structural repertoire of the human V kappa domain. *EMBO J.* **14**, 4628-4638.

Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhammer, E. L. L. & Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256**, 813-817.

Ulrich, H. D., Patten, P. A., Yang, P. L., Romesberg, F. E. & Schultz, P. G. (1995). Expression studies of catalytic antibodies. *Proc. Natl Acad. Sci. USA*, **92**, 11907-11911.

van Dijk, K. W., Mortari, F., Kirkham, P. M., Schroeder, H. W. & Milner, E. C. (1993). The human immunoglobulin VH7 gene family consists of a small, polymorphic group of six to eight gene segments dispersed throughout the VH locus. *Eur. J. Immunol.* **23**, 832-839.

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotech.* **14**, 309-314.

Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucl. Acids Res.* **22**, 5600-5607.

Wagner, S. D. & Neuberger, M. S. (1996). Somatic hypermutation of immunoglobulin genes. *Annu. Rev. Immunol.* **14**, 441-457.

Wall, J. G. & Plückthun, A. (1999). The hierarchy of mutations influencing the folding of antibody domains in *Escherichia coli*. *Protein Eng.* **12**, 605-611.

Williams, S. C., Frippiat, J. P., Tomlinson, I. M., Ignatovich, O., Lefranc, M. P. & Winter, G. (1996). Sequence and evolution of the human germline V-lambda repertoire. *J. Mol. Biol.* **264**, 220-232.

Winter, G. (1998). Synthetic human antibodies and a strategy for protein engineering. *FEBS Letters*, **430**, 92-94.

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433-455.

Wörn, A. & Plückthun, A. (1998). An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Letters*, **427**, 357-361.

Wörn, A. & Plückthun, A. (1999). Different equilibrium stability behavior of scFv fragments: Identification, classification and improvement by protein engineering. *Biochemistry*, **38**, 8739-8750.

Wörn, A., Auf der Maur, A., Escher, D., Honegger, A., Barberis, A. & Plückthun, A. (2000). Correlation between in vitro stability and in vivo performance

of anti-GCN4 intrabodies as cytoplasmic inhibitors. *J. Biol. Chem.* in the press.
Wu, H., Beuerlein, G., Nie, Y., Smith, H., Lee, B. A., Hensler, M., Huse, W. D. & Watkins, J. D. (1998). Stepwise in vitro affinity maturation of Vitaxin, an alpha(v)beta(3)-specific humanized Mab. *Proc. Natl Acad. Sci. USA*, **95**, 6037-6042.
Wu, T. T., Johnson, G. & Kabat, E. A. (1993). Length distribution of CDRH3 in antibodies. *Proteins: Struct. Funct. Genet.* **16**, 1-7.
Yang, W. P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R. & Barbas, C. F. (1995). CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.* **25**, 392-403.
Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.
Yélamos, J., Klix, N., Goyenechea, B., Lozano, F., Chui, Y. L., González, Fernández A., Pannell, R., Neuberger, M. S. & Milstein, C. (1995). Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature*, **376**, 225-229.

Edited by I. A. Wilson

(Received 12 August 1999; received in revised form 3 December 1999; accepted 6 December 1999)

http://www.academicpress.com/jmb

Supplementary material for this paper is available from JMB Online.